Universidad
Carlos III de Madrid
www.uc3m.es

TESIS DOCTORAL

Essays in
Empirical Industrial Organization

Autor:

Luis Aguiar

Director:

Philippe Gagnepain

DEPARTAMENTO DE ECONOMÍA

Getafe, Diciembre de 2013

Universidad
Carlos III de Madrid
www.uc3m.es

TESIS DOCTORAL

Essays in Empirical Industrial Organization

Autor: *Luis Aguiar*

Director: **Philippe Gagnepain**

Firma del Tribunal Calificador:

Firma

Presidente:

Vocal:

Secretario:

Calificación:

Getafe, de de

Universidad Carlos III de Madrid

Doctoral Thesis

Essays in Empirical Industrial Organization

Author:
Luis Aguiar

Supervisor:
Dr. Philippe Gagnepain

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

in the

Department of Economics

December 2013

“Be not the slave of your own past. Plunge into the sublime seas, dive deep and swim far, so you shall come back with self-respect, with new power, with an advanced experience that shall explain and overlook the old.”

Ralph Waldo Emerson

Abstract

My PhD thesis consists of three chapters in Empirical Industrial Organization. The first two chapters focus on the relationship between firm performance and specific public policies. In particular, we analyze the cases of cooperative research and development (R&D) in the European Union and the regulation of public transports in France. The third chapter focuses on copyright protection in the digital era and analyzes the relationship between legal and illegal consumption of digital music.

The first chapter, entitled *European Cooperative R&D and Firm Performance*, focuses on the impact of participation in research joint ventures as part of the European Union Framework Programmes on firms' economic performance. These programmes are the main financial tools used by the European Union to support cooperative R&D activities in the EU. Unlike previous empirical studies, this chapter suggests that their impact on firms' competitiveness is significant. We analyze industry-oriented research joint ventures supported by the Fifth European Framework Programme between 1998 and 2002. A key feature of this Programme is that funding is available to the firms based on social and economic concerns instead of pure performance criteria, which guarantees that financial support is not granted conditional on technological opportunities. This allows us to identify the causal effect of the programme on firms' performance using the funding available to the firms in their respective industries as a source of exogenous variation in the decision to participate in the programme. Our results suggest that participation in large research projects raises labor productivity by at least 35 percent and profit margin by up to 8 percentage points.

The objective of the second chapter, entitled *Knowledge Spillovers in Cost-Reduction Incentives*, is to identify and measure the relevance of knowledge spillovers in the French urban transportation industry, where most regulated transportation networks are operated by firms that belong to the same company. We build and estimate a structural cost regulation model under incomplete information where the service is regulated by an authority and is provided by a single operator that may be owned by a larger company. We identify the knowledge spillovers which arise for some operators being linked to a same group, and see how they influence the firms' decisions of exerting effort in order to reduce their operating costs. Our model provides us with estimates of the operators' inefficiencies, the effort of the managers and the knowledge spillovers. Our results show that knowledge spillovers are indeed relevant for the existing industrial groups present in the French urban transport industry. Simulation exercises provide evidence of significant reductions in total operating cost following the enlargement of industrial groups and mergers between existing groups.

In the third chapter, entitled *Digital Music Consumption on the Internet: Evidence from Clickstream Data*, we analyze the behavior of digital music consumers on the Internet. Using

clickstream data on a panel of more than 16,000 European consumers, we estimate the effects of illegal downloading and legal streaming on the legal purchases of digital music. Our results suggest that Internet users do not view illegal downloading as a substitute to legal digital music. Although positive and significant, our estimated elasticities are essentially zero: a 10% increase in clicks on illegal downloading websites leads to a 0.2% increase in clicks on legal purchases websites. Online music streaming services are found to have a somewhat larger (but still small) effect on the purchases of digital sound recordings, suggesting complementarities between these two modes of music consumption. According to our results, a 10% increase in clicks on legal streaming websites lead to up to a 0.7% increase in clicks on legal digital purchases websites. We also find important cross country difference in these effects.

Resumen

Mi tesis doctoral consta de tres capítulos en Organización Industrial Empírica. Los dos primeros capítulos se centran en la relación entre el rendimiento de la empresa y determinadas políticas pubicas. En particular, se analizan los casos de cooperación en Investigación y Desarrollo (I+D) en la Unión Europea y la regulación de transporte público en Francia. El tercer capítulo se centra en el derecho de autor (copyright) en la era digital y analiza la relación entre consumo legal e ilegal de música digital.

El primer capítulo, titulado *European Cooperative R&D and Firm Performance*, se centra en el impacto de los Programas Marco de la Unión Europea (European Union Framework Programmes) sobre los resultados económicos de las empresas participantes. Estos Programas tienen como objetivo fomentar la cooperación en I+D, subvencionando a empresas para que participen en proyectos de cooperación en I+D (Research Joint Ventures). A diferencia de estudios empíricos previos, este capítulo sugiere que su impacto sobre la competitividad de las empresas participantes es significante. Se analizan proyectos de cooperación en I+D orientados hacia la industria y financiados por el Quinto Progrma Marco de la Unión Europea entre 1998 y 2002. Una característica clave de este programa es que los fondos de financiación disponibles para las empresas se basan en criterios sociales y económicos en lugar de criterios de rendimiento puro, lo que garantiza que la ayuda financiera no se otorga en función de oportunidades tecnológicas. Esto nos permite identificar el efecto causal del programa sobre los resultados de las empresas usando los fondos disponibles para las empresas en sus respectivas industrias como fuente de variación exógena en la decisión de participar en el programa. Nuestros resultados sugieren que la participación en grandes proyectos de investigación aumenta la productividad del trabajo en al menos un 35 por ciento y el margen de beneficio en hasta 8 puntos porcentuales.

El objetivo del segundo capítulo, titulado *Knowledge Spillovers in Cost-Reduction Incentives*, es identificar y medir la relevancia de externalidades de conocimiento (knowledge spillovers) en la industria de transporte urbano francés, donde la mayoría de las redes de transporte reguladas son operadas por empresas que pertenecen a un mismo grupo empresarial. Se construye y se estima un modelo estructural de regulación bajo información incompleta, donde el servicio es regulado por una autoridad y es proporcionado por un único operador que puede pertenecer a una empresa mayor. Se identifican las externalidades de conocimiento que surgen del hecho que algunos operadores están vinculados a un mismo grupo empresarial, y se analiza cómo influyen al esfuerzo de las empresas para reducir sus costes operativo. El modelo nos proporciona estimaciones de las ineficiencias de los operadores, del esfuerzo de los directivos y de las externalidades de conocimiento. Los resultados muestran que las externalidades de conocimiento son relevantes para los grupos industriales presentes en el sector del transporte urbano francés. Ejercicios de simulación proporcionan evidencia de una

reducción significativa en el coste operativo total tras ampliar los grupos industriales y tras fusiones entre los grupos existentes.

En el tercer capítulo, titulado *Digital Music Consumption on the Internet: Evidence from Clickstream Data*, se analiza el comportamiento de los consumidores de música digital en Internet. Se hace uso de datos de visitas (clickstream) que permite seguir el comportamiento de más de 16.000 consumidores europeos en Internet y se estiman los efectos de las descargas ilegales y del streaming legal sobre las compras legales de música digital. Nuestros resultados sugieren que los usuarios de Internet no ven las descargas ilegales como un sustituto a la música digital legal. Aunque positivas y significativas, las elasticidades estimadas son esencialmente cero: un aumento del 10 % de los clics en los sitios web de descargas ilegales lleva a un incremento del 0,2% de los clics en sitios web de compra legal. Servicios de streaming de música en línea tienen un efecto algo mayor en las compras de música digital, lo que sugiere una complementariedad entre estos dos modos de consumo de música . De acuerdo con nuestros resultados, un aumento del 10 % de los clics en los sitios web de streaming legales conduce a un aumento de hasta el 0,7 % de los clics en sitios web de compra legal. También encontramos importantes diferencias entre países en estos efectos.

Acknowledgements

It's hard to believe I'm writing this. Hard to believe this day has come. It's been a long journey, and I must thank many people for helping me complete it. I would like to thank my advisor, Philippe Gagnepain, as well as all the professors who have helped me, one way or another. Many thanks to Matilde Machado, Bruno Cassiman, Georges Siotis, Natalia Fabra, Diego Moreno and Ulrich Wagner. I would also like to thank Bertin Martens for his kindness and support during the past year.

Thank you Dani. We've shared many years together, from the first day of the PhD until the end of it. Always in the same office, too! Thank you for putting up with me, and for your invaluable advice on any project/problem set/question/doubt I may have had. You've made economics clearer to me so many times. Thank you for being such a good friend, and of course for all the *cafés & pulguitas*. Thank you for believing in me more than I did sometimes. I owe you big time.[1] Juan, Becu, pibe, tritri… Gracias por las tapas, las cañas, las charlas, los debates, los chistes, el apoyo en los momentos más difíciles y todo lo demás. Gracias Luisfa, por siempre estar dispuesto a salir y charlar un rato. Gracias por las risas y por hacer que no me sintiera tan sólo en este viaje tan tumultuoso que ha sido el doctorado. Thanks to Lucila, Dolo and Joaco for their support and words of encouragement. Thanks to María for the crazy preparations of TA sessions. Many thanks to all professors and fellow PhD students at Universidad Carlos III de Madrid who helped me during all these years!

Finalemente, gracias a Diana. Por estar aquí. Por apoyarme y creer en mí, siempre.

[1] I also owe you for letting you down for breakfast so many mornings, I know you hated me for that…

Contents

List of Figures

List of Tables

Pour ma mère, qui m'a toujours soutenu. Merci d'être là et de nous montrer l'exemple à suivre chaque jour.

Pour mon père, qui m'a toujours poussé.

Pour Diego, en espérant qu'il puisse être aussi fier de moi que je ne le suis de lui.

Para mis abuelos, quienes me enseñaron el valor de la educación.

Para Diana, porque sin ti nada valdría la pena.

Chapter 1

European Cooperative R&D and Firm Performance: Evidence Based on Funding Differences in Key Actions

1.1 Introduction

Research and development (R&D) investments are flawed by two important characteristics that make their equilibrium levels less than socially desirable in a freely competitive market. First, the knowledge generated by a firm's R&D effort is non-rival: To the extend that this knowledge cannot be kept secret, its use by a firm does not preclude its use by another. Second, R&D is characterized by spillovers: A firm investing in R&D usually imposes a positive externality on the other firms which can appropriate the results of this investment.[1] This will lead firms to under invest and therefore to an under-provision of R&D investment in the economy.

Along with the establishment of an intellectual property system, two types of public policies are generally used to reduce this market failure. First, direct subsidies can be offered to firms. By modifying the marginal return of R&D investments, they encourage firms to invest more than they would in a free market equilibrium.[2] A second policy consists in encouraging firms to collaborate in R&D activities in order to partially internalize the externality they impose on other firms. In this paper we focus on this last type of policy. More specifically, we focus on the core instrument used by the European Union to support European cooperative

[1] See De Bondt (1997) for a review.

[2] The government can also intervene and encourage R&D investments through tax incentives.

R&D activities, the European Union Framework Programmes (EU-FPs in the remainder of the paper).

The main objective of European policies toward research joint ventures in the beginning of the 1980's was to fight the relative decline in the international competitiveness of high technology sectors.[3] Started in 1984, the first Framework Programme came in response to a situation where individual R&D activities were uncoordinated and required a large number of Council decisions (Georghiou, 2001).[4] The EU-FPs are the main financial tools used by the EU Commission to support cooperative R&D activities, and the EU participation in the coordination and financing of RJVs has been increasing until today.[5] Due to the large amount of public funds raised by the different EU-FPs, it is crucial to have a clearer idea about their effect and the outcomes they generate. To help in accomplishing this task, the present paper analyzes the effect of participation in the Fifth Framework Programme (EU-FP5 in the remainder of the paper), which was allocated a total budget of 14.96 billion euros over the 1998-2002 period; this amounts to almost 2% of the total intramural R&D expenditures generated by the EU 27 countries over the same period (Source: Eurostat). More specifically, we focus on its effects on two firm level performance measures, labor productivity and profitability.

The predecessors of the EU-FP5 mostly aimed at stimulating the transnational collaboration in research, particularly between industry and universities (European Commission, 2000, 2001). The important role of these types of partners in shaping projects' objectives indicates that these were primarily oriented towards explorative research rather than market exploitation of research results.[6] In other words, most of the research carried before the EU-FP5 did not intend to develop specific products and processes on its own, which makes it "pre-competitive". Pre-competitive research concerns R&D for which commercial possibilities remain five to ten years in the future (Luukkonen, 1998). This characteristic has largely explained the poor direct effects on the economic results of participants found in previous studies (Benfratello and Sembenelli, 2002; Barajas et al., 2011).

[3]Other factors specific to the European Community (EC) also influenced the need for these policies. For instance, there were large differences between the many country members in terms of industrial and technological capabilities. Some members also had an already well established policy infrastructure for Science and Technology while others totally lacked such infrastructures. Finally, there was no appropriate legal framework and institutions at the EC level for supporting a consistent technology policy. In 1981, these considerations led the European Commission to establish the pilot ESPRIT program with the endorsement of the twelve largest European producers of electronics (Hagedoorn et al., 2000).

[4]Also at that time arose the formal expression of the policy rationale for the Community action in the field of research and technological cooperation. This is embedded in the principle of subsidiarity, which states that support should come where the scale or cost of cooperation was beyond that affordable by a single country, where complementarity in national work could achieve results for the whole Community, and where research contributes to development of the common market, laws and standards, or to the unification of European science and technology (Georghiou, 2001).

[5]The 1st, 2nd, 3rd, 4th and 5th EU-FPs were allocated 3.75, 5.4, 6.6, 13.2, and 14.96 billion euros, respectively (Artis and Nixson, 2001).

[6]Exploration is understood as "the pursuit of knowledge, of things that might come to be known," and exploitation as "the use and development of things already known" (Levinthal and March, 1993).

Instead, the EU-FP5 includes an important thematic programme, namely the User-friendly Information Society (IST in the remainder of the paper) programme, which includes projects that remain mainly industry-driven (Fisher et al., 2009). As opposed to participants coming from research and academic communities, industry partners are more likely in this case to be driven by motives to commercially exploit rather than explore a given technology. Projects involving mainly industry partners, even if not targeted to the development of a particular marketable product or service, are consequently associated with objectives that are closer to the market. The mechanism through which performance could be enhanced by participating in the programme is not explicitly modeled here, but we have in mind that cooperative R&D agreements are part of an innovation activity that provides access to external know-how and hence leads to gains in performance. This know-how is expected to have a more direct impact on performance when collaboration is more market-oriented.[7] We argue that focusing more specifically on the IST programme allows us to identify a significant effect of participation in the EU-FP5 on firms' performance.

The main econometric challenge of our study arises from the fact that participation in the EU-FP5 is not random. Participation is the result of a selection process involving decisions from both participants and the European Commission. Participants must first decide to joint an RJV and elaborate a proposal. The Commission then decides whether to fund (part of) the project. Hence, showing that participating firms perform better than non-participating ones is not sufficient to prove a positive impact of programme participation. This self-selection problem is crucial and recurrent when estimating the impact of government sponsored R&D. Not taking it into account would severely bias the results (Klette et al., 2000). To get rid of this self-selection effect, we use a two step estimation method where we first estimate a selection equation. For this purpose we need at least one exogenous variable that provides randomness in the participation decision but that is otherwise unrelated to firms' performance.

We use the funding available to the firms in their respective industries as a source of exogenous variation in the decision to participate in the programme. We expect this variable to be an important determinant of the participation status of each firm, since the higher the funding available the higher the willingness to participate and/or the higher the likelihood that the project is accepted and funded. A relevant concern is that the European Commission might allocate its support partly in line with technological opportunities, which could in turn differ across industries and affect firms' performance. We take advantage of a key feature of the EU-FP5 which is that funding is available to the firms through key actions based on social and economic concerns instead of pure performance criteria. According to the European Commission, "the idea of the key actions is precisely to bring together the contributions of

[7] In general, the empirical literature corroborates that a more market-oriented collaboration is more likely to bring along positive economic effects (Belderbos et al., 2004; Cincera et al., 2003).

specialists from very differing scientific fields, together with industrial researchers, users, and political and economic decision-makers."[8] This would lead to the development of projects that allow, for instance, people to choose, order, and pay electronically in complete safety, or to design a system "to provide users with a full range of transport-related information such as parking availability, traffic jams, recommended routes, public transport, and so on."[9] Since funding is not motivated by performance, it can be used as a tool to solve the selection issue. This specific variable has to our knowledge never been used in the analysis of these programmes nor in the context of RJV studies. It has however been used to identify the effects of specific contracts on firms' R&D investments (Lichtenberg, 1988) and in the context of R&D subsidies (Wallsten, 2000; Gelabert et al., 2009). In a recent paper, Einiö (2012) follows a similar approach and uses geographic differences in R&D support allocation as an instrumental variable to assess the effects of government R&D subsidies on company performance.

We evaluate the effect of participation in the programme on performance across two important dimensions. First, R&D collaboration remains an activity with long-term objectives, and this is a crucial feature that needs to be taken into account. In our analysis, we make sure to identify the long-term effect of participation in the programme on the economic performance of firms. In particular, our database allows us to consider lags of up to 4 years after the start of each project. Second, we account for the heterogeneity in the projects' size to better understand how participation may affect firms' performance. If RJV size and diversity increase knowledge complementarity and therefore spillover effects among participants, it would also lead to a larger increase in R&D efforts (Sakakibara, 2001). Moreover, if higher R&D expenditures increase a firm's absorptive capacity and learning capability, participation in an RJV should as well increase a firm's R&D effort (Cohen and Levinthal, 1989). Thus, participation in large RJVs will lead to larger impacts in terms of productivity and profit gains since they induce participants to put more effort in them.

Our results suggest that the long-term effect of participation is an increase in labor productivity by, at least, 35 percent. We also find a positive long-term effect of participation on the profit margin, with increases of up to 8 percentage points. The large magnitude of our estimates will be put into perspective. In particular, our results will be interpreted as the average impact of the programme for those firms induced to participate as a result of the change in the funding available to them (the "marginal" participants).

The remainder of the paper is organized as follows. Section 1.2 summarizes the relevant literature on the subject. It presents the results of the main empirical studies on the effects of participation in the EU-FPs and relates them to the programmes' characteristics. Section

[8]See http://ec.europa.eu/research/rtdinf21/en/key/why.html.
[9]See http://ec.europa.eu/research/rtdinf21/en/key/07.html.

1.3 presents the EU-FP5 in more detail as well as the IST programme. The empirical strategy for identifying the causal effect of participation in the IST programme on economic performance is presented in section 1.5, while section 1.4 presents the data and the different variables used in the estimation. Section 1.6 is devoted to the presentation and discussion of our results. Finally, section 1.7 draws some policy implications and concludes.

1.2 Related literature

Our paper shares features with two important categories of empirical studies on R&D collaboration. It is first related to the empirical analysis of the determinants of RJV formation and participation. As an important part of this rather thin literature, Hernán et al. (2003) analyze the determinants of participation in European RJVs and find that sectorial R&D intensity, industry concentration, firm size, technological spillovers, and past RJV participation positively influence the probability of forming RJVs. Marín and Siotis (2008) extend this analysis by exploiting the differences in institutional design of two European collaboration programmes (EUREKA and the EU-FPs) and find that past experience in the EU-FPs is an important factor explaining participation. For the case of US RJVs, Röller et al. (2007) take asymmetries in firms' size into account and show that these are important determinants of participation. They find that larger firms are less willing to share their economic knowledge with smaller rivals.

Second, our work relates to empirical studies analyzing the effect of cooperation on firm's economic performance, such as productivity or profits.[10] Even though this literature has resulted in quite mixed results, it has supported the existence of a positive relationship between close-to-the-market R&D cooperation and economic performance.

An early work analyzing the effect of RJV participation on firm economic performance is the one by Siebert (1996). Analyzing 314 US joint ventures, he shows that cooperation has no direct impact on profit margin, but he finds that the effect of R&D intensity on the profit margin is larger for cooperating than for non-cooperating firms. In a very influential paper analyzing the effects of collaboration in Europe, Belderbos et al. (2004) study the impact of cooperation on Dutch firms' productivity. They differentiate between the type of R&D partner (competitors, suppliers, customers, and universities and research institutes) and find that supplier and competitor cooperation has a significant impact on labor productivity growth. They do not, however, find any significant impact of cooperation with universities or research institutes on labor productivity, highlighting the importance of market orientation

[10] Another part of the literature has analyzed the effects of R&D cooperation on *innovative* performance, like sales of innovative products or patenting activity (Branstetter and Sakakibara, 1998, 2002; Dekker and Kleinknecht, 2008; Czarnitzki et al., 2007).

for the effects of collaboration on economic performance. Cincera et al. (2003) take the view that cooperation in R&D gives access to external know-how and use it to explain performance at the firm level. Using data on R&D and productivity for Belgian firms, they find that on top of own R&D expenditures, international R&D cooperation significantly increases a firm's productivity growth. Just as in Belderbos et al. (2004), they put forward the fact that firms may benefit differently from different types of cooperation and find that the main benefits come from international cooperation with customers, suppliers or other companies, which reflects more applied international cooperative activities. Their results therefore give further evidence on the positive relationship between the degree of market orientation of the cooperation and its impact on economic performance.

The empirical literature concerning the effects of collaboration taking place in the EU-FPs has shown rather disappointing results, mainly explained by the pre-competitiveness nature of the projects. Benfratello and Sembenelli (2002) carry an analysis to evaluate the impact of European collaboration programs on participating firms' productivity. They study the impact of two different programs, EUREKA and the (3rd and 4th) EU-FPs in the 1992-1996 period. They find that firms participating in EUREKA have experienced a significant improvement in their performance measures, while firms participating in RJVs under the EU-FP scheme do not show any significant change in performance. They attribute this result to the fundamental differences between the two programmes. The EUREKA programme has a decentralized funding source where research projects are proposed and defined by the participants themselves. It therefore shows a bottom-up structure which has much more market-oriented projects, as opposed to the top-down structure of the EU-FPs and their pre-competitive projects. In a recent study, Barajas et al. (2011) analyze the impact of participation in the EU-FP on the productivity of Spanish manufacturing firms between 1995 and 2005.[11] They show that participation has a positive impact on firms' technological capabilities, which in turn have an effect on firms' labor productivity. In other words, they do not find a direct effect of participation on economic performance, but they find an indirect effect through the generation of new knowledge.

The characteristics of the EU-FPs (pre-competitiveness, participation of universities and research institutes) have lead their impact to be mainly set on firms' technological development and capacity. Luukkonen (1998) shows that their main impact has indeed been intangible effects, such as learning new skills or creating new network relations.[12] Other studies have also found these impacts to differ with firms' characteristics, and in particular with respect to size. Fisher et al. (2009) analyze the relationship between participation in the EU-FP5 and EU-FP6 and the innovative activity of firms using data from the Community Innovation Survey and a large database composed from other sources. They find that, as opposed

[11] Their analysis therefore covers parts of EU-FP4, all of EU-FP5 and part of EU-FP6.
[12] Skills refer to the technical and scientific skills rather than to the social skills needed in collaboration.

to large companies, small and medium enterprises demonstrate more economically-driven objectives (innovation, commercialization and market-related) and generally join a project looking for complementary resources to achieve a specific objective that will typically be a new or improved product/service or process. This translates into more positive results in terms of innovation. They also notice that, due to their limited size and resource level, SMEs will engage in a small number of cooperative agreements each of which will be important for their immediate survival and growth. For these type of firms, the funding provided by the commission is therefore crucial. Finally, a relevant finding of their study is the positive effect on both product and process innovation for first-time participants in the EU-FPs.

The next section is now devoted to a more detailed presentation of the EU-FP5 on which we will concentrate our empirical analysis.

1.3 The EU-FP5, the IST Programme, and Key Actions

Since 1984, research and innovation activities from the EU are bundled into the EU-FPs. These have been the main financial tools with which the EU supports R&D activities covering almost all scientific disciplines. Six EU-FPs have already been completed and the seventh has started in 2007.[13] The aim of these EU-FPs is to support and encourage European research, but the detailed objectives of each programme vary from one funding period to another. All of the RJVs that are formed under this programme are eligible for an EU subsidy, which varies according to the nature of the project.

The EU-FP5 comprises several thematic programmes, which are themselves decomposed into a total of 23 Key Actions. The thematic programmes are "Quality of Life and Management of Living Resources", "User-friendly Information Society (IST)", "Competitive and Sustainable Growth", "Energy, Environment and Sustainable Development", and "Nuclear Energy". In this paper we focus on the IST programme. Two main reasons motivated our choice. First, with a budget of 3.6 billion euros, the IST programme represents the lion's share of the EU-FP5 in terms of budget allocation. The second reason is tightly linked to the objectives set by the commission in the design of the EU-FPs' projects. The pre-competitiveness of a project, as argued above, is recurrently mentioned in the empirical literature as being the reason for the poor economic effects observed on the firms participating in the EU-FPs. Our view is that the cooperation taking place in the projects of the IST programme have an impact on economic performance through the sharing of knowledge and the learning of new skills. Given their more industry-oriented nature, these projects are more likely to be driven by motives to commercially exploit rather than explore a given technology. We therefore

[13] The seven EU-FPs cover the periods 1984-1987, 1987-1991, 1990-1994, 1994-1998, 1998-2002, 2002-2006, and 2007-2013.

believe the relationship between access to knowledge and firm performance to be of a more direct nature in the IST programme.

The IST programme contains four Key Actions: Key Action 1 is called *Systems and services for the citizen*; it aims at improving information and communications technologies in a wide variety of domains such as health, education, culture, social services, the needs of elderly and handicapped people, the environment, transportation and leisure. An example is the project directed by Nokia which leads to the development of a portable terminal combining mobile telephony and PDA (Personal Digital Assistant) technology. According to the European Commission's webpage, "The system is designed to provide users with a full range of transport-related information such as parking availability, traffic jams, recommended routes, public transport, and so on. Six towns in Finland, Sweden, the United Kingdom, the Netherlands, France and Germany have hosted tests for this innovation, in conjunction with several major European telecommunications firms, car manufacturers and GIS (geographical information system) providers."[14] Key Action 2 is denoted *New methods of work and electronic commerce*; its objective is to develop telework and electronic commerce and investigate an in-depth reorganisation of social relations and labour legislation, both for business and for individuals. An example is the SEMPER Project (Secure Electronic Marketplace for Europe). As described on the European Commission's webpage, it "has developed one of the first operational architectures tailored for commerce on the Internet. Using the web, consumers can access a database of catalogues of goods and services, and fill in order forms on their computer screens. Payment is by credit card, using the SET protocol (Secure Electronic Transaction), or by an e-cash smart card."[15] Key Action 3 is related to *Multimedia content and tools*. The European Commission highlights the importance of multimedia technologies as they are "opening new ways of mastering information, acquiring knowledge, and transferring know-how available to a broad public." As an example, their website presents the project "SAVIE (Support Action for Videoconferences In Education) which has produced several training modules which have permitted teachers to prepare and produce lessons that are adapted to the new teaching tools."[16] Finally, Key Action 4 is called *Essential technologies and infrastructures*; it focuses on essential components involving micro-electronics and software engineering, which deal with processing, storing and transmitting information in many types of products and services. A project example, also taken from the European Comission's webpage, is the one of "ASML, which has become a lead player in the domain of photolithography - a strategic technology for printing the integrated circuits found in micro-processors. ASML is developing a technology of scan photolithography, which is revolutionising productivity and the cost of printing integrated circuits one tenth of a micron

[14] See http://ec.europa.eu/research/rtdinf21/en/key/07.html.
[15] See http://ec.europa.eu/research/rtdinf21/en/key/08.html.
[16] See http://ec.europa.eu/research/rtdinf21/en/key/09.html.

insize."[17]

The design of Key Actions is an important novelty of the EU-FP5 in the history of the EU-FPs. They aim at identifying socio-economic stakes and concentrating research funds in order to develop research activities that are organized around key issues. Thus, promoting research focused on performance for its own sake is not relevant here. This is a very important property since it suggests that the funds invested in the EU-FP5 by the European Commission are not targeting specific industries based on their performance. At the time of identifying the causal effect of participation in the IST programme on firms' performance, the funds made available by the programme in each industry is an excellent source of exogenous variation. Note, however, that we still expect participation in the IST programme to help firms to potentially improve their performance as we picture cooperative R&D agreements as part of an innovation activity that provides access to external know-how and hence leads to gains in performance. This know-how is expected to have a more direct impact on performance when collaboration is more market-oriented.

The European Commission does not itself undertake or participate in the EU-FP projects. Its role is to offer financial or other support to private and public research bodies, and companies and institutions wishing to embark on a research project. Each year throughout the period of the EU-FP5, the commission publishes so-called workprogrammes that contain different calls for proposals that describe the objectives planned (Zobel, 1999). The proposal of a project must then be submitted in response to these specific calls. This means that unsolicited project proposals are not allowed and the project's content must correspond to the objectives set out by the commission. Also, several eligibility criteria must be satisfied by the different partners involved in the project. One of them is that the project must involve at least two legal entities (e.g. individuals, industrial and commercial firms including SMEs, universities, research bodies, technology dissemination bodies) independent of each other and established in two different Member States or in a Member State and an associated country.[18] The financial contribution from the Commission consists in the reimbursement of a set percentage of the participants' eligible expenses, although sometimes flat-rate contributions are made. In order to be reimbursed by the Commission, participants must identify and report their eligible expenses by submitting interim and final statements. In particular, the expenses must be necessary for the action in question, provided for in the contract, actually incurred and recorded in the accounts. Finally, it is important to note that participants cannot establish intellectual property rights over their discoveries: all research must be shared among partners.

[17]See http://ec.europa.eu/research/rtdinf21/en/key/10.html.

[18]This means that entities established outside the EU and international organizations can also participate.

1.4 Data

Conducting a study on the impact of participation in the IST programme requires a database that contains both information on the different projects included in the programme and on the economic performance of firms for a period long enough to capture the long term effects of collaboration. The empirical analysis will therefore be carried out using a database constructed from two different sources. The data from the IST projects is taken from the Community Research and Development Information Service (CORDIS) web page, where a total number of 2522 projects is available.[19] The second source of information is the one about the participating firms. Once the information about each project is recovered, we can look at each participating firm individually in order to obtain firm-level data. This latter task will be done using AMADEUS (Analyse MAjor Databases from EUropean Sources), a database produced by BUREAU VAN DIJK, a specialist provider of firm-level data. Firms participating in the projects recovered from the CORDIS web page are therefore linked to the AMADEUS database in order to retrieve their relevant information. The AMADEUS database contains balance sheet information on the top 250,000 firms in Europe, while the CORDIS database provides information on each project, i.e. its description, its reference, the starting and ending dates of the project, its status and its acronym, the contract type offered to the participants, the cost of the project as well as the funding provided by the European Commission. The name of the coordinator of the project and of the participating firms are given as well.

We were able to retrieve 961 firms that participated in at least one FP5 RJV from AMADEUS. Table 1.1 gives the different number of RJVs the firms participate in and shows how some firms were often involved in more than one project. In our analysis, we decided to focus on the firms that participate in one project only. This corresponds to a total of 620 firms participating in 466 projects. After cleaning the dataset, we end up with a total of 379 participants that correspond to 315 projects.[20]

Table 1.2 presents some average values for the projects included in the database. The column *All Projects* represents all the projects we could recover from the CORDIS webpage for the EU-FP5 (2359 projects)[21], while the column *Single Part* contains the projects in which only single participants (in our data set, that is) are involved (466 projects). The last column *Sample* contains information on the projects that correspond to the participating firms present in our final sample (315 projects). The projects that we are able to analyze

[19] All the projects' fact sheets are available at http://cordis.europa.eu/ist/projects/projects.htm.

[20] We realize that the final number of participating firms in our sample is rather low. To correct for the fact that AMADEUS contains larger and potentially more productive firms, we construct different control groups to deal with this potential bias. They are presented in greater detail below.

[21] Due to some technical restrictions related to our data collection procedure, we were not able to recover the information on all of the 2522 projects available on the CORDIS webpage.

seem to be larger in terms of number of participants and cost. Unless otherwise stated, the next tables will present statistics of the projects included in our final sample.

Table 1.3 reports the characteristics of the projects in our database according to their starting dates. The vast majority of the projects were initiated between 2000 and 2002, and only a few in 2003. Table 1.4 provides summary statistics on the number of participants by project, showing that projects are more or less evenly distributed, with a higher proportion incorporating 6 to 10 participants. The duration is on average lower when projects have few participants (0 to 3) and the cost of each RJV is increasing with the number of participating firms. Regarding the projects' costs, table 1.5 reveals that the majority of RJVs have costs between 0 and 6 millions euros, with a peak for the ones with costs between 1 and 3 millions. We can also observe that both the number of participants and their diversity in terms of industry increase with the cost of the project. When carrying our analysis on the effects of participation taking project's size into account, we will define a large RJV as being one with a total budget of more than 2.8 million euros, which is the median value of the distribution of the projects' cost in our sample. Table 1.6 presents projects' characteristics when classified according to our definition of size. Again, note that large projects not only involve more participants, but that they involve more participants coming from different industries (defined at the 4-digit level). Large projects are therefore more diverse than small ones in terms of participants' industry of origin.

An important problem one has to deal with when evaluating the impact of government-sponsored R&D is the one of selection bias since it is hard to think of RJV participation as being randomly assigned or decided. This inevitably creates a potentially important bias in the estimated impact parameters. Table 1.7 provides us with a glimpse of this potential problem by reporting summary statistics on some variables for both the participants and non-participating firms in AMADEUS for 1999. Participants have significantly larger figures for most of the variables considered, confirming the fact that the programme selects larger firms for participation. Further evidence of this fact is given in the left panel of figure 1.1, which shows the distributions of the log transformation of sales for both participants and firms contained in AMADEUS. The participants' distribution is similar to the one of the outsiders, only shifted to the right.

To perform our empirical test, we use three different samples. The first one is composed of the participating firms and of non-participating firms randomly picked from AMADEUS. After cleaning the data, we are left with 2134 observations for participants and 6638 for the selected non-participants over the years 1997 to 2006. This sample is referred to as the RANDOM sample throughout the text. An alternative control group is constructed by selecting non-participating firms from AMADEUS so as to replicate the cross-tabulation of participants by country and industry. After cleaning the data, we are left with 2134 observations for participants and 3531 for the selected non-participants over the years 1997 to 2006. In our estimations, we call this sample IC-Rep (for Industry Country Replication). Last, we use a third control group constructed so as to replicate the distribution of the sales variables of the participants for 1999 (i.e. before the start of any project). The final kernel density estimates of the control group for 1999 are presented in figure 1.1. After cleaning the data, we are left with 2134 observations for participants and 3726 for the selected non-participants over the years 1997 to 2006. We call this sample Sales-Rep in our estimations.

A potential concern is that firms belonging to our control group may be involved in other RJVs. Another important programme under which pan-European RJVs have been formed in the last two decades is the EUREKA programme, another initiative aimed at enhancing cross-border technological cooperation. In order to further support the validity of our samples, we would therefore like to verify that the non-participating firms present in our control groups are not participating in other R&D collaboration programmes such as EUREKA. We were able to do so for some of the firms in our database, as we were given access to information on the French firms that participated in the EUREKA programme during the years 1998 to 2005. We were therefore able to check whether French non-participants from our control groups had participated in the EUREKA programme during this same period.[22] Although our control groups are not only composed of French firms, the latter still represent a non-negligible share of the non-participants with 16.4 percent, 19.3 percent and 23.6 percent in our three different samples. The results showed that only 5 firms did participate in EUREKA during he same period for one of the samples, while for the other two control groups, only 1 and 4 firms participated. This means that more than 95 percent of the French firms in our samples have not participated in the EUREKA programme.

Finally, the repetitive nature of the Framework Programmes rises a concern as well. Indeed, if firms currently participating in the EU-FP5 have been involved in previous Framework Programmes, identifying the sole effects of a participation in the EU-FP5 on firms' performance becomes tricky. This concern is specially relevant since the previous literature has found that many participants tend to repeat their participation in consecutive editions of the programme (Hernán et al., 2003; Barajas et al., 2011). We do not have any information on the EU-FP4; however, we have data on the EU-FP6 which allows us to check whether participants in the IST programme of the EU-FP5 are also involved in the IST programme of the EU-FP6. The result of this exercise revealed that out of the 379 participants present in our final sample, less than 14 percent (51 firms) took part in the IST programme of the EU-FP6. It suggests that our sample includes a small share of firms that are prone to repeat the experience. As for the non-participating firms, only a very small fraction (less than 0.01 percent for each of the three different samples) turned out to be participating in the EU-FP6, giving further support to the validity of our control groups.

[22]We are grateful to Aminata Sissoko for allowing us to do so.

1.5 Empirical strategy

We provide an empirical test of the effect of participation in the IST programme on the firms' economic performance. Let $P_{it} = 1$ be the event of firm i participating in a project at time t and let y_{it} be the measure of firm i's performance. Denote by y^0_{it} and y^1_{it} the performance of firm i at time t when it does not and when it does participate in EU-FP5 respectively. Hence we write

$$y_{it} = \begin{cases} y^1_{it} & \text{if } P_{it} = 1 \\ y^0_{it} & \text{if } P_{it} = 0. \end{cases}$$

Equivalently, y_{it} can be expressed as

$$y_{it} = y^0_{it} + (y^1_{it} - y^0_{it})P_{it}. \tag{1.1}$$

We want to identify the effect of participation at time t on the firm's performance y_{it}. This effect can be expressed as $\Delta_{it} \equiv y^1_{it} - y^0_{it}$. It measures the difference between the observed performance of participant i and the performance it would have reached had it not participated in the project. Since the counterfactual outcome y^0_{it} can never be observed for a participating firm, Δ_{it} cannot be computed directly and needs to be estimated. If we consider a constant treatment effect, i.e. $\Delta_{it} = y^1_{it} - y^0_{it} = \delta$, we can rewrite (1.1) as

$$y_{it} = \alpha + \delta \cdot P_{it} + \varepsilon_{it}, \tag{1.2}$$

where α is a constant and ε_{it} is an error term. A direct approach to circumvent the missing counterfactual problem is to replace the missing counterfactual outcome by the mean performance of the non-participating firms. This would be a simple treatment-control comparison (TCC) estimator as it mimics the analysis in an experimental setting. The estimator of δ, $\widehat{\delta}^{TCC}$, would then be the mean difference in performance between participants and non-participants.

A simple treatment-control comparison in the form of equation (1.2) is most likely to yield inconsistent estimates. As mentioned above, $\widehat{\delta}^{TCC}$ will suffer from a selection bias since it is hard to think of participation in the programme as being random. Selection bias comes from the existence of firms' characteristics (be they observable or not) that are correlated with participation in the programme. To the extend that the programme attracts bigger and more productive firms, we have to deal with a positive selection bias. We therefore also control for observable characteristics x that affect both the decision to participate (treatment) and the productivity of the firm (outcome). Doing so leads to the following specification:

$$y_{it} = x'_{it}\beta + \delta \cdot \mathrm{PART}_{it} + \varepsilon_{it}. \tag{1.3}$$

Estimation of δ from equation (1.3) allows to control at least for selection on observable characteristics (all included in the vector x) such as firm size, capital intensity, absorptive capacity, industry concentration as well as country, industry and time fixed effects.

To the extend that firms self-select in the programme based on some *observable* characteristics, the above estimation strategy allows us to solve for the selection problem. It is, however, most likely that firms decide on participation based on *unobservable* characteristics included in ε_{it} as well, in which case the endogeneity problem will remain and estimators will still be inconsistent. We can, for example, think of firms as having heterogeneous "managerial" or "innovative" ability that may influence their decision to participate in an RJV. Participation decisions (from the firms or from the programmes' organization) may also be based on past outcomes of y_{it}. Klette et al. (2000) give an example from the study of Klette and Moen (1999) in which the Norwegian government was supporting large firms facing severe problems when the IT industry was restructured towards the end of the 1980's. In this case, we would have that $\mathrm{COV}(\varepsilon_{is}, \mathrm{PART}_{it}) \neq 0$ for $s < t$, leading to inconsistent estimation results of the impact of participation.

When identification is jeopardized because the participation (or treatment) variable is endogenous, a standard solution is to look for a variable that generates some exogenous variation in the participation decision of firms, which would allow to mimic a randomly assigned treatment. Finding such a variable is not easy, as it amounts to finding a variable that simultaneously determines the participation decision of the firms and does not appear as a determinant of the outcome variable y_{it}.

Two-steps estimation

The empirical strategy then consists in two steps. In the first one, we specify an equation explaining the participation decision. In particular, we assume that the probability that firm i will engage in an RJV of the IST programme is given by

$$\Pr(\mathrm{PART}_{it} = 1 \mid \mathbf{z}) = \mathrm{F}(z'_{it}\gamma). \tag{1.4}$$

The variables in the vector x in (1.3) are a subset of the variables in z. That is, at least one element of z (call it z_1) is unique and is a non-trivial determinant of PART_{it}. Hence z_1 is a variable correlated with the endogenous dummy variable PART_{it} but that has no direct effect on the outcome y_{it} (it only has an effect through PART_{it}). We will specify $F(\cdot)$ to be the cumulative distribution function of the logistic distribution.

The methodology consists in estimating (1.4) using pre-determined observations to explain programme participation. Notice that this amounts to using pre-programme firm characteristics as instruments for the endogenous variable PART_{it}. Given that the IST programme starts in the year 2000, we want to prevent firms' characteristics from being affected by the programme. In essence, the first step of our estimation strategy therefore tries to define the profile of a typical participating firm right before the start of the IST programme. For this purpose, we use observations for years 1997 to 1999 to estimate equation (1.4) and obtain an estimate of $\hat{\gamma}$. Notice that in order to do so we use as a dependent variable a dummy indicating whether the firm will be a participant in the programme. We then construct the predicted probabilities of participation using the subsequent years of data and our estimate $\hat{\gamma}$ to obtain $\widehat{PART}_{it} = F(Z_{it}\widehat{\gamma})$. Notice that although we use past firms' characteristics as instruments, the current firms' characteristics variables are used to construct the predicted probabilities of participation. The second step of our strategy consists in using these predicted values to identify the impact of participation estimating equation (1.3) using years 2000 to 2006.

The variables

Our econometric specification requires the construction of a set of variables that measure or proxy the determinants of participation in the IST programme as well as the determinants of our outcome variables (labor productivity and profit margin). The performance measures that will be considered are labor productivity, measured as added value per employee, and profit margin, measured as the profit (before taxation) over the operating revenue.

The most important explanatory variable is the one that we use as a source of exogenous variation to explain participation. Our approach is based on the idea that differences in available funding across industries induce variation in the likelihood of participating in the programme. Indeed, the participation in a project is the result of two decisions. The initial decision comes from the firms, which must choose whether to apply or not for funding. Conditional on the result of this first decision, the European Commission then decides whether to fund the project or not.[23] The budget dedicated to the funding of RJVs is therefore likely to be correlated with the participation decision of firms for at least two (non-exclusive) reasons. First, firms will be more willing to participate if they know that more funds are available. Second, a project is more likely to be accepted if the commission has more funding to offer.

[23] Unfortunately, we only observe the accepted projects in our dataset. The firms that applied for funding but were denied can therefore not be distinguished from those that did not apply.

An important concern is that the Commission might allocate its support partly in line with technological opportunities, i.e., the projects that are selected are those which involve industries that perform badly. As explained in detail above, the exogeneity of our available funding variable relies on the creation of the key actions in the EU-FP5. Indeed, Edith Cresson, the then European Commissioner in charge of research and innovation stated that "*We are moving from research based on performance for its own sake to research which focuses on the social and economic problems which face society today*".[24] Thus, the objective underlying the Fifth Framework Programme differs radically from that of its predecessors.

Since any industry could potentially be represented in any of the Key Actions, the latter provide exogenous variation in the availability of funding in each industry. Optimally, we would like to observe the part of the budget of each KA that goes to each industry so as to build a measure of available budget at the industry level. Since we do not observe these shares, we need to build our available funding variable based on the awarded funds in each industry:

$$AvailableFunding_j = \sum_{k \in KAs} \sum_{RJV_r^k} d_{jr}^k \cdot \text{Funding}_r^k$$

where d_{jr}^k is a dummy equal to 1 if a firm from industry j participates in RJV r in key action k, and Funding_r^k is the funding received by RJV r in key action k from the EU-FP5.

To compute our two steps estimation procedure, we use additional explanatory variables. First, we introduce a measure of firms' size to take into account the existent asymmetries across firms. As noted by Röller et al. (2007), differences in firms' sizes reflect differences in efficiency. This variable may also have an important effect on participation in case specific fixed costs for the creation of an RJV exist. For example, large firms would be able to spread these costs more easily across a larger volume of sales and would therefore be more willing to participate in the programme. Another measure of size that must be taken into account is the *relative* size of a firm within its industry. As noted by Hernán et al. (2003), relative size may matter if RJVs are used as a vehicle for pursuing "technology watch", i.e. to monitor innovative activity in their segment. As they point out, the largest firms (which are also the technology leaders), have most to lose from the emergence of new, technologically advanced rivals (see also Laredo (1998)). This measure is proxied by the introduction of a variable measuring market share, calculated as firm size over industry size, both measured by the amount of sales.[25]

[24] See "A turning point for community research", RTD Info 21, p.3 at http://ec.europa.eu/research/rtdinf21/en/edito.html.

[25] This index is constructed over the entire AMADEUS database at the four-digit industry level.

R&D expenses are an important determinant of firm's participation in the programme as they are a good measure of a firm's "absorptive capacity". This idea was first introduced by Cohen and Levinthal (1989), who argue that external knowledge is more effective for the innovation process when the firm engages in own R&D. Performing R&D would therefore increase a firm's value of cooperation and increase its willingness to participate in such agreements.[26] One main shortcoming of our dataset, however, is the unavailability of R&D expenses at the firm or even at the industry level. Although R&D expenses are not explicitly reported in AMADEUS, they are, in most countries, booked under intangible assets. In order to partially overcome this availability problem, we use the ratio of intangible fixed assets over employees (in logarithm) as a proxy for the intensity of the firm's innovative activity. We realize that this variable also contains information on patents, copyrights, trademarks and other similar items and may therefore not give a perfect measure of R&D intensity. This variable is however likely to be highly correlated with a firm's absorptive capacity, increasing the likelihood of participation in an RJV.

Industry concentration has an ambiguous effect on the incentives to participate in R&D collaboration. On the one hand, a highly concentrated industry can facilitate the identification of suitable partners and spillovers to non-participants are limited because of their reduced number. Also, an RJV may well be created in order to weaken competition or increase the market power of its participants. In all of these cases, more concentration would increase the incentives for firms to participate in RJVs. On the other hand, one could also expect a negative impact of concentration on the likelihood of RJV formation since strict limits are imposed by competition policy on collaborative projects in concentrated industries.[27] To construct a measure of industry concentration, we include the Herfindähl-Hirschman Index (HHI) for each four-digit sector present in our sample.[28] The HHI is defined as:

$$\mathrm{HHI}_j = \sum_{i=1}^{n} (MarketShare_{i,j})^2.$$

Further control variables include a set of 2-digit industry dummies as well as country dummies and the ratio of tangible fixed assets over employees (in logs) as a proxy of physical capital intensity.

With these covariates properly defined, we can now respectively rewrite equations (1.4) and (1.3) as

[26] See Cassiman and Veugelers (2002) for a discussion on the effects of absorptive capacity on the probability of cooperating in R&D.

[27] An example is the EU's block exemption which automatically allows ventures between firms that collectively represent less than 25 percent of the relevant anti-trust market but requires authorization for values above that threshold.

[28] Similarly to our market share measure, this index is constructed over the entire AMADEUS database.

$$
\begin{aligned}
\Pr(\mathrm{PART}_{it} = 1 \mid \mathbf{z}) =& \mathrm{F}(z'_{it}\gamma) \\
=& \mathrm{F}\Bigg(\gamma_0 + \gamma_1 \log(Employees)_{it} + \gamma_2 \log\left(\frac{FixedAssets}{Employees}\right)_{it} \\
&+\gamma_3 \log\left(\frac{IntangibleAssets}{Employees}\right)_{it} + \gamma_4 \mathrm{HHI}_{jt} + \gamma_5 \mathrm{MktShare}_{it} \\
&+\gamma_6 \log(AvailableFunding)_j \\
&+\sum_{p=1}^{P} \gamma_{7p} Country_{ip} + \sum_{g=1}^{J} \gamma_{8g} Industry_{ig} + \sum_{s=98}^{99} \gamma_{9s} ds_t\Bigg)
\end{aligned} \tag{1.5}
$$

and

$$
y_{it} = x'_{it}\beta + \delta \cdot \mathrm{PART}_{it} + \varepsilon_{it}, \tag{1.6}
$$

where x'_{it} contains all the variables of z'_{it} excluding $AvailableFunding$. We estimate Equation (1.5) with a logit procedure and obtain $\widehat{\mathrm{PART}}_{it}$; in a second step, $\widehat{\mathrm{PART}}_{it}$ is used as an instrument in (6). In our regressions, we bootstrap the standard errors to account for the fact $\widehat{\mathrm{PART}}_{it}$ is an estimated variable. Since the residuals are also likely to be correlated within industries, our calculation of standard errors controls for this correlation by clustering at the four-digit industry level.

1.6 Results

We now present the results of our estimations. We first discuss the results concerning the determinants of participation in the programme and then turn to the effects of the programme on economic performance.

Determinants of participation in the IST programme

Table 1.8 presents the results of the logit estimation (1.5) of the determinants of participation in the IST programme, controlling for residual correlation among observations from the same industry. For each of our different samples (Random, IC-Rep and Sales-Rep), we present two alternative specifications in order to assess whether the results are sensitive to the inclusion of the intangible assets intensity as a proxy for R&D intensity in determining participation. The results appear to be robust to the inclusion of this variable as the other coefficients are not significantly affected.

Our main attention is set on the parameter associated to the variable $AvailableFunding$. The coefficient turns out to be positive and strongly significant in both specifications for our three different samples, corroborating the fact that the available funding is indeed an important predictor of participation in the programme.[29] As explained above, two possible non-exclusive explanations can explain this result. One is the fact that firms are more willing to participate (i.e. to apply for a subsidy) when the available funding is larger. Another possibility is that, all else equal, firms that are willing to participate (i.e. that already applied for participation) are more likely to be accepted for a subsidy if the funding is larger. Although we are not able to identify which is the true mechanism driving this correlation with the data at hand, either one of them serves our purpose by confirming the relevance of our exogenous variable.

The coefficient associated with firm size is positive and highly significant in the IC-Rep and Random samples.[30] As already noted by Hernán et al. (2003), several non-exclusive explanations can explain this finding. First, controlling for industry concentration, large firms may have a preference to collaborate with other large firms in order to maximize the internalization of spillovers (see Röller et al. (2007) for a theoretical model). Second, it may reflect the existence of large fixed costs associated with RJV formation (for example large administrative and negotiation efforts necessary to reach agreements with partners, the establishment of specific facilities). Third, for projects involving partners from different and complementary industries, a preference to cooperate with a larger partner may simply reflect a preference to cooperate with a more efficient partner. Finally, the positive coefficient associated with firm size may also be the result of a certain exogenous preference for large firms on the part of the EU-FP5 organization.

The coefficient associated with firm market share, a measure of the firm's relative size, is positive and significant in both specifications for our three samples. This results corroborates the "technology watch" explanation presented above, according to which relatively large firms in an industry (i.e. leaders) have an incentive to participate in the programme to monitor the innovative activity in their segment. Indeed, technological leaders have a lot to lose from the emergence of technologically advanced rivals.

[29]Table 1.8 presents the χ^2 statistics for the coefficient on the $AvailableFunding$ variable in each sample and specification. The values of the test are all above 50 and strongly reject the null of the available funding not affecting participation in the programme. Since we are explaining future participation in the programme with data from 1997 to 1999, it could nevertheless be that the available funding does not predict participation that well in the years following 2000. To check for that possibility we used the data from 2000 to 2006 to run regressions of our participation variable on the predicted values $\widehat{PART}_{it}$ and all the variables included in the vector x in (1.3). The results present positive and highly significant coefficients on our predicted participation variable.

[30]This coefficient is not significant in the case of the Sales-Rep sample given that we constructed the latter replicating the participants' sales, a variable highly correlated with firms' number of employees.

Although significant in only two of the samples, the HHI variable shows a positive impact on the probability of participation[31], indicating that firms coming from more concentrated (or less fragmented) industries are more likely to participate. As argued above, this result is consistent with the fact that firms find it easier to identify suitable partners in such industries. Also, the latter provides greater scope for the internalization of spillovers.

The fixed assets intensity, a measure of capital intensity, is a positive predictor of participation, but turns out to be significant in the Random sample, and only when the intangible fixed assets intensity is not included as a regressor, see specification (1). When the latter is included, its corresponding coefficient is positive and significant, showing the important correlation between the fixed assets and intangible fixed assets variables.

Finally, the coefficient on the intangible assets intensity variable shows up to be positive in our three samples, but only marginally significant for the Random sample. Although the sign is the one expected, we can therefore not affirm that R&D activities proxied by the intangible fixed assets are a significant determinant of programme participation.

Impact of the IST programme on economic performance

In the second step of our procedure, we replace the participation dummy by the predicted value $\widehat{\text{PART}}_{it}$ obtained from the participation equation. The impact of participation on the firms' performance is then estimated using observations from years 2000 to 2006.

Before turning to the discussion of the results, it is important to recall the interpretation that must be given to our estimates. To the extent that the treatment effects are heterogeneous among different firms, our strategy allows us to estimate the average treatment effect for the firms whose treatment status (participant or not) is affected by changes in the variable $AvailableFunding$. In this case we are therefore not able to identify the average treatment effect on all the treated, but only for the *marginal* participants. For this effect to be identified, an additional monotonicity assumption still needs to be met, which says that while the exogenous variable might have no effect on some firms, all of those who are affected in their participation decision must be affected in the same way, see Imbens and Angrist (1994). Our results should therefore be interpreted as the average impact of the programme for those firms induced to participate as a result of the change in the funding available to them.

Tables 1.9 and 1.10 present the estimation results for Equation (6). In each of the tables and for the three different samples used, the columns (OLS) report the OLS estimates, while columns (TS1) and (TS2) show the results of our two-stage procedure. The OLS estimates

[31] Just as in the case of firm size for the Sales-Rep sample (see footnote 30), the non-significance of the concentration index for the IC-Rep sample is most probably due to the way we constructed the latter (i.e. replicating the cross tabulation of participants by country and industry).

suggest a positive effect of participation on the labor productivity, whereas the effect on profit margin is mainly non significant. Since OLS ignores the endogeneity of participation in the programme, these estimates are likely to be biased if selection into the programme is based on unobservable characteristics. Columns (TS1) and (TS2) present the results of estimating Equation (6) correcting for the endogeneity of participation. We find that the average effect of participation on labor productivity is positive and significant in our three samples. Firms engaging in an RJV enjoy an average increase in labor productivity of about 25 to 34 percent. At the same time, Table 1.10 suggests that the effect of participation on profit margin is nil.

As R&D collaboration is an activity with long term objectives, we also attempt to identify lagged effects of participation on firms' performance over time.[32] As the mean duration of a project in the sample is 27 months, we may expect the effects of a project to appear at least 2 years after its start. Hence, we re-estimate Equation (6) as follows:

$$y_{it+\tau} = x'_{it}\beta + \delta_\tau \cdot \mathrm{PART}_{it} + \varepsilon_{it+\tau},$$

where the dependent variable $y_{it+\tau}$ refers to the $(t+\tau)$th period after the starting year of the observed project. The coefficient δ_τ must then be interpreted as the average impact of programme participation on economic performance, starting τ years after entering the project. Comparing the coefficients δ_τ for different values of τ will therefore help to see the evolution and distribution of the impact of participation over time. Tables 1.11 and 1.12 report the δ_τ coefficients (for $\tau = 0, \ldots, 4$) for each of our estimations. Each line therefore shows a point estimate resulting from a different regression estimation.[33]

We first discuss the results in table 1.11, which refer to labor productivity as a measure of economic performance. Except for the Random sample which presents slight drops in the point estimates, we observe an increase in the magnitude of the δ_τ coefficients when τ increases. This suggests that, overall, the effects of participation in the programme on labor productivity are significant and should be measured from a long-term perspective. On average, participation leads to a significant increase in labor productivity of about 30 percent to 38 percent three to four years after starting the project. Turning to the effects on the profit margin (table 1.12), our results only show significant impacts in the Sales-Rep sample. In particular, the point estimates indicate that participation leads to an increase of 4 to 5 percentage points in the profit margin. As in the case of labor productivity, the

[32]The need to measure the long term impact of participation in EU-FP has already been noted by most empirical analysis (Dekker and Kleinknecht, 2008; Benfratello and Sembenelli, 2002; Barajas et al., 2011).

[33]The first line of each table therefore reports the coefficients on the participation dummies from tables 1.9 and 1.10 respectively (i.e. when $\tau = 0$).

evolution of the coefficients shows that the effect of participation becomes significant several years after the start of a project.

Analysis by RJV size

The size of the project may influence the magnitude of the impact of firms' participation in the EU-FP5 on their performance; indeed, the size of a project is related to the RJV diversity, which in turn affects the R&D effort of participants and increases knowledge sharing among partners.

Sakakibara (2001) highlights three possible channels through which the diversity of an RJV could positively affect the R&D effort of participants: The first channel relates to the spillover effect of a firm's own R&D on others' R&D productivity. Assuming that RJV size and diversity increase knowledge complementarity among participants, a higher degree of knowledge complementarity implies a larger spillover effect, and would also lead to a larger increase in R&D efforts. Second, RJVs provide firms with new learning opportunities. If higher R&D expenditures increase a firm's learning capability, and assuming that better learning opportunities arise when the size and the diversity of the RJV increases, participation in a large RJV will lead to larger R&D efforts by each RJV participant. Finally, when cooperative R&D reduces firms' marginal costs of production, the resulting increase in competition (and decrease in profits) will lead to a lower level of R&D effort. Large RJV are more likely to involve participants coming from different industries, reducing the risk of an increase in competition which, in turn, decreases the likelihood of a reduction in the R&D effort.

We therefore test whether heterogeneity in projects' size translates into heterogeneous effects in terms of economic performance. To do so, we separate large projects from small projects in the following specification:

$$y_{it} = x'_{it}\beta + \delta^L \cdot \mathrm{PART}_{Lit} + \delta^S \cdot \mathrm{PART}_{Sit} + \varepsilon_{it}, \tag{1.7}$$

where $\mathrm{PARTS}_{Lit} = 1$ if firm i participates in a large (L) project and $\mathrm{PARTS}_{Sit} = 1$ if firm i participates in a small (S) project. The vector x includes the same covariates as in equation (6). Since participation in a project of a given size is again endogenous, we use our two-step approach to estimate the effect of participation in the two different kinds of RJVs. Our strategy is similar to the one we followed in section 1.5 with the difference that we explain now participation in the two types of projects (large versus small): Each firm i can therefore choose among three alternatives, participate in an large project, participate in a small project or participate in neither of them (i.e. stay out of the programme). We define the dependent variable PART_{ic} to take value 1 if project c is ever chosen, where $c \in \{\text{Large}, \text{Small}, \text{Out}\}$. We therefore assume that the probability that firm i chooses project c is given by

$$\Pr(\text{PART}_{ict} = 1 \mid \mathbf{z}) = \text{F}(z_{it}'\gamma_c), \tag{1.8}$$

where $\text{F}(\cdot)$ is now the multinomial logistic cumulative distribution function. The estimation strategy is identical to the one we pursue in section 1.5. In order to mitigate endogeneity issues, we use pre-programme observations ($t = 1997, 1998, 1999$) to explain the participation choices in the several categories specified above. We then use observations for years 2000 to 2006 to predict participation decisions using the results from the first step and subsequently use the predicted values to identify the impact of participation in the two types of RJVs by estimating equation (1.7) with the 2000-2006 period. Again, we control for residual correlation among observations from the same industry and bootstrap the standard errors in our regressions.

Table 1.13 reports the results of the multilogit estimation (1.8). Regardless of the sample used and the methodology (including or not the intangible assets intensity), the coefficient on the $AvailableFunding$ variable is positive and strongly significant for both types of RJVs (although larger for big projects). The available funding is therefore a relevant determinant of participation in the programme, irrespectively of the size of the project considered. Almost all of the remaining explanatory variables have the same sign as in our simple logit model. The coefficient on firm size is quite similar for large and small projects, and is always positively related to participation.[34] The coefficient on the HHI variable again shows a positive relationship between industry concentration and participation in the two different types of projects. The results for the Random sample are, however, not significant anymore for either type of RJV.[35]

Capital intensity is a positive predictor of participation, and always a more important one for larger projects. Although not significant, it even shows to be negative for small projects in two of our samples. On the contrary, the intangible assets intensity is a stronger positive predictor for participation in smaller project, although only significantly so in the Random sample. Finally, an interesting result appears for the variable referring to firms' relative size. As in the simple logit case, it is again positively related to participation, but is now significant only for larger RJVs. This again corroborates our "technology watch" interpretation and further suggests that for a leading firm, monitoring the innovative activity in its sector is more relevant when the projects are of an important size. These kinds of projects are indeed the ones where competing firms could gain the most and possibly take the technological lead.

[34] Again, the coefficients appear non significant for the Sales-Rep sample, see footnote 30.

[35] For the same reason presented in footnote 31, they are neither significant for the IC-Rep sample.

In the second step of our procedure, we replace the participation dummies (one for each type of project) by the predicted values obtained with the information from the first step. Tables 1.14 and 1.15 present the estimation results for equation (1.7) when residual correlation among observations from the same industry is accounted for. The OLS estimates suggest a positive effect of participation on the labor productivity, for both small and large projects. This effect is however greater for large projects. No significant effect is obtained if profit margin is the explained variable. The IV estimation results suggest that the average effect on productivity over the years following the start of the projects is positive and significant in our three samples. The average gains in labor productivity come mainly from the larger projects and range from 35 percent to close to 50 percent. The results regarding the profit margin show no significant impact of participation in either type of project.

Instead of an average effect, we may as well evaluate how the impact of participation is distributed across years after the start of the project. Table 1.16 focuses on labor productivity and suggests that, in the RANDOM sample, the effect of participation in small projects is insignificant from 3 years after the start of the project, while it is larger and significant for the large projects. The same pattern is observed for the remaining two samples, although the impact for the small projects is never significant. According to these estimates, the long-term effect of participation in a large project from the IST programme is an increase in labor productivity in about 37 percent to up to 60 percent.

Table 1.17 presents the results when performance is measured by the profit margin. The results reveal a positive and significant impact from participating in large RJVs, while participation in small projects leads to negative and significant effects on the profit margin. Participation in large projects leads to increases in as much as 8 percentage points, while participation in small projects leads to decreases to up to 6 percentage points in one of our samples.

The results obtained when taking the size of the projects into account give support to the underlining mechanism that we expect to be at work: Cooperative R&D agreements are part of an innovation activity that provides access to external know-how which leads to gains in performance. In this respect, large projects provide many advantages that may lead to increases in participants' R&D efforts compared to small projects. In particular, their more diverse composition in terms of industry origin offers RJV participants better learning opportunities and increases spillover productivity.

1.7 Discussion and conclusion

In this paper we analyze the effects of R&D collaboration within the EU-FP5 on firms' economic performance. Previous literature has shown that participation in RJVs supported by the EU-FPs has had little direct relevant impact on firms' economic outcomes, a fact mainly explained by the pre-competitiveness of the programme. By concentrating on the IST programme, we focus our analysis on the projects that involve more market-oriented collaboration, and which are therefore more likely to result in direct positive economic effects. We also account for the fact that R&D collaboration remains an activity with long-term objectives and therefore identify the long-term effect of participation in the programme.

As a mean to address the self-selection effect of participation, we follow a two-step method and use the funding available to the firms as an exogenous variable to provide randomness in the firms' participation status. Our results show that the long-term effects of participation is an increase in labor productivity by, on average, almost 40 percent. Taking projects' size into account, this increase appears to be mainly driven by gains in the large projects, as we find that entering a large RJV raises labor productivity by up to 60 percent. We also find a positive effect of participation on the profit margin, with increases of 4 to 5 percentage points. These positive effects are again the result of the important impact of participation in the larger projects, which leads to gains of up to 8 percentage points.

The large magnitude of our estimates has to be put into perspective. Indeed, our results should be interpreted as the average impact of the programme for those firms induced to participate as a result of the change in the funding available to them. Our results should therefore not necessarily be taken as evidence of the aggregate effectiveness of the EU-FP5, but as the average effect on the "marginal" participants. Though we are not able to identify these particular firms, our results on the determinants of participation may give us a hint about their characteristics. We found absolute firm size to be an important determinant of participation, pointing to the fact that RJVs involve large fixed costs. The "marginal" participants, whose participation in the programme is more dependent on the funding available and received, are most likely to be smaller, first-time participants. This is in line with the results of Fisher et al. (2009) which found first-time participants and medium-sized companies to benefit the most from participation in the EU-FP5 and EU-FP6 in terms of innovation. We see participation in the IST programme as a way of obtaining access to new knowledge and resources which in turn positively affect economic performance.

It is also important to note that participation in the IST programme actually involves two simultaneous actions, namely cooperation with other firms or institutions (i.e. the formation of an RJV) and the granting of a subsidy to help financing the project pursued by the RJV. We are unfortunately not able to disentangle these two effects separately, and can *a priori* only identify a joint effect of both cooperation and subsidy granting. Our results on the impact of participation by project size indicate that cooperation within an RJV and the sharing of know-how is a crucial factor to explain the gains in performance from the IST programme. This, however, is not informative on the direct effect of the funds received by the RJV. One may argue that our results would be consistent with a scenario in which RJV are beneficial (the mere fact of cooperating with other firms) but the subsidy itself is not, meaning that the gains from cooperation would have been obtained regardless of the granting of the subsidy. We stress, however, that some firms (in particular small or financially constrained firms) would not be able to participate in an RJV if there was no subsidy, and that our results show that the benefits of participation can be very substantive for these specific firms.

Raising the available funding for the small first-time participants would encourage them to participate in projects that would benefit them greatly. This could be accomplished, for instance, by covering a substantial part of their fixed costs, such as the administrative costs for the project's proposal or for the research project itself. In any case, and as suggested by Barajas et al. (2011), policy makers should take these costs into account and distinguish between firms with previous experience in cooperative projects and other firms. In particular, participation in large projects would lead to important gains in competitiveness.

Table 1.1: Number of RJVs per firm

Number of RJVs	Number of firms	Per cent	Cumul.
1	620	64.52	64.52
2	140	14.57	79.08
3	64	6.66	85.74
4	34	3.54	89.28
5	21	2.19	91.47
6	14	1.46	92.92
7	8	0.83	93.76
8	7	0.73	94.48
9	7	0.73	95.21
10 or more	46	4.79	100.00

Table 1.2: Mean statistics by project

Variable	All Projects	Single Part	Sample
Nb of Participants	7.00	8.58	8.77
Duration (in months)	27.04	27.93	27.40
Cost (thousand €)	2376.54	2999.21	3002.23
Funding (thousand €)	1380.21	1663.37	1638.60
Nb of Projects	2359	466	315

Table 1.3: Projects' characteristics by starting year

Starting Year	Number of RJVs	Number of participants	Duration in months	Cost in thousand €	Funding in thousand €
2000	96 (30.5 %)	8.23	27.55	3096.99	1683.81
2001	108 (34.3 %)	8.69	27.62	2761.45	1502.80
2002	104 (33.0 %)	9.33	26.68	3240.63	1786.52
2003	7 (2.22 %)	9.29	32.57	1875.71	916.29
All	315	8.77	27.40	3002.23	1638.60

TABLE 1.4: Projects' characteristics by number of participants

Number of participants	Number of RJVs	Duration in months	Cost in thousand €	Funding in thousand €
3 or less	14 (4.4 %)	17.21	631.56	444.35
4 to 5	36 (11.4 %)	28.19	2429.06	1345.53
6 to 7	89 (28.2 %)	27.94	2657.61	1476.00
8 to 10	105 (33.3 %)	27.12	2874.11	1581.15
11 to 15	48 (15.2 %)	29.25	4163.96	2167.86
16 or more	23 (7.3 %)	27.65	4836.39	2611.22
All	315	27.40	3002.23	1638.60

TABLE 1.5: Projects' characteristics by cost

Cost in millions	Number of RJVs	Duration in months	Number of participants	Nb of different industries	Funding in thousand €
0 to 1	53 (16.8%)	18.49	6.68	1.55	430.35
1 to 3	122 (38.7%)	28.35	8.13	1.81	1191.41
3 to 6	111 (35.2%)	29.91	9.33	2.59	2145.41
6 to 8	21 (6.7%)	30.48	13.00	3.43	3394.76
more than 8	8 (2.5%)	29.00	13.63	4.38	4821.25
All	315	27.40	8.77	2.21	1638.60

TABLE 1.6: Projects' characteristics by size†

Size of RJV	Number of RJVs	Number of participants	Duration in months	Cost in thousand €	Funding in thousand €	Nb of different industries
Small	162 (51.4%)	7.73	24.83	1523.66	909.75	1.73
Large	153 (48.6%)	9.88	30.12	4567.78	2410.33	2.72
All	315	8.77	27.40	3002.23	1638.60	2.21

† An RJV of small size is defined as one with a total cost of less than 2.8 million €.

Table 1.7: Comparison of participants and Amadeus for 1999

	Participants			Amadeus		
	Mean	Median	N	Mean	Median	N
Sales	1,617,717.1	61,192.0	560	76,461.5	12,648.0	91475
Employees	7,833.5	441.0	522	482.4	87.0	102273
Fixed Assets	1,295,842.3	15,400.5	598	60,009.1	2,112.0	146753
Intangible Fixed Assets	147,837.6	381.5	582	5,413.6	2.0	140781
Labor Productivity	317.9	63.2	380	141.7**	49.0	71872
Costs of Employees	366,136.9	20,507.0	508	14,431.1	1,882.0	106638
Mean Wage	49.6	42.5	448	57.5***	31.1	91759
Profit Margin	5.8	4.9	560	5.0***	3.0	124496
Gross Profit Margin	40.1	35.9	140	69.0***	18.3	42718

** Cannot reject the null hypothesis (equality of the means) in a two-tailed t-test at the 5% level between the participants and the corresponding control group in each column.

*** Cannot reject the null hypothesis (equality of the means) in a two-tailed t-test at the 10% level between the participants and the corresponding control group in each column.

Table 1.8: First stage estimation results (logit)†

Sample	Random		IC-Rep		Sales-Rep	
	(1)	(2)	(1)	(2)	(1)	(2)
Constant	-6.795***	-6.153***	-8.635***	-8.357***	-6.917***	-9.288***
	(1.49)	(1.51)	(2.10)	(2.16)	(1.64)	(2.05)
log(Employees)	0.427***	0.451***	0.501***	0.502***	0.012	0.013
	(0.09)	(0.09)	(0.08)	(0.08)	(0.07)	(0.07)
log(Fixed Assets Intensity)	0.201***	0.134*	0.104	0.085	0.055	0.037
	(0.07)	(0.08)	(0.07)	(0.07)	(0.08)	(0.08)
log(Intang Assets Intensity)		0.120**		0.039		0.033
		(0.05)		(0.05)		(0.05)
log(Available Funding)	0.507***	0.499***	0.459***	0.455***	0.515***	0.513***
	(0.06)	(0.06)	(0.06)	(0.06)	(0.07)	(0.07)
Market Share	6.229***	6.044***	6.621***	6.452***	3.123**	3.047**
	(2.18)	(2.32)	(2.35)	(2.38)	(1.61)	(1.60)
HHI	2.271**	2.303**	1.216	1.181	4.238***	4.237***
	(1.12)	(1.11)	(0.82)	(0.82)	(1.20)	(1.19)
Industry dummies	YES	YES	YES	YES	YES	YES
Country dummies	YES	YES	YES	YES	YES	YES
Time dummies	YES	YES	YES	YES	YES	YES
χ^2 test on						
log(Available Funding)	60.34	60.96	52.27	51.91	74.67	74.41
Pseudo-R^2	0.436	0.443	0.291	0.292	0.370	0.370
Number of obs.	1667	1667	1545	1545	1362	1362

† The dependent variable is equal to 1 for participants and 0 for non-participants. Standard errors in parenthesis and clustered at the four-digit industry level.

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

Table 1.9: Second stage estimation results: Labor Productivity†

Sample	Random			IC-Rep			Sales-Rep		
	(OLS)	(TS1)	(TS2)	(OLS)	(TS1)	(TS2)	(OLS)	(TS1)	(TS2)
Constant	1.992***	2.172***	2.176***	1.214*	1.661***	1.228*	2.109***	2.120***	2.226***
	(0.65)	(0.59)	(0.61)	(0.68)	(0.52)	(0.70)	(0.28)	(0.30)	(0.29)
log(Employees)	-0.148***	-0.165***	-0.162***	-0.118***	-0.137***	-0.136***	-0.103***	-0.104***	-0.106***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)
log(Fixed Assets Intensity)	0.266***	0.268***	0.263***	0.253***	0.254***	0.251***	0.268***	0.280***	0.267***
	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)
log(Intang Assets Intensity)	0.015**		0.010	0.006		0.004	0.019**		0.018**
	(0.01)		(0.01)	(0.01)		(0.01)	(0.01)		(0.01)
Market Share	2.322***	2.498***	2.456***	0.955	0.939	0.925	0.708	0.781	0.726
	(0.65)	(0.68)	(0.68)	(0.90)	(0.95)	(0.94)	(0.51)	(0.51)	(0.50)
HHI	0.232	0.159	0.186	0.091	-0.024	-0.014	-0.055	-0.162	-0.148
	(0.19)	(0.19)	(0.19)	(0.15)	(0.17)	(0.16)	(0.16)	(0.18)	(0.18)
PARTS	0.181***	0.370***	0.317***	0.161***	0.353***	0.337***	0.092**	0.276**	0.247*
	(0.04)	(0.12)	(0.12)	(0.03)	(0.11)	(0.11)	(0.05)	(0.14)	(0.14)
Industry dummies	YES	YES	YES	YES	YES	YES	YES	YES	YES
Country dummies	YES	YES	YES	YES	YES	YES	YES	YES	YES
Time dummies	YES	YES	YES	YES	YES	YES	YES	YES	YES
Adjusted-R^2	0.712	0.710	0.711	0.502	0.501	0.501	0.579	0.577	0.580
Number of obs.	7105	7105	7105	5879	5879	5879	4498	4498	4498

† The dependent variable is the logarithm of labor productivity. Standard errors in parenthesis are bootstrapped and clustered at the four-digit industry level. In specifications (OLS), the variable PARTS is a simple dummy equal to 1 if the firm participates in EU-FP (Pooled OLS is used). The variable PARTS for columns (TS1) and (TS2) corresponds to the predicted values of the first stage logit estimations (1) and (2) in table 1.8 respectively.

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

TABLE 1.10: Second stage estimation results: Profit Margin†

Sample	RANDOM			IC-REP			SALES-REP		
	(OLS)	(TS1)	(TS2)	(OLS)	(TS1)	(TS2)	(OLS)	(TS1)	(TS2)
Constant	-0.029	-0.035	-0.041	-0.146**	0.088**	-0.140*	0.022	0.046	0.030
	(0.08)	(0.08)	(0.09)	(0.07)	(0.04)	(0.07)	(0.04)	(0.03)	(0.04)
log(Employees)	0.003	0.003*	0.003*	0.002	0.001	0.001	0.005**	0.004**	0.005**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
log(Fixed Assets Intensity)	0.009***	0.008***	0.009***	0.006**	0.003	0.006**	0.006***	0.004*	0.006***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
log(Intang Assets Intensity)	-0.002**		-0.002**	-0.004***		-0.004***	-0.002**		-0.003***
	(0.00)		(0.00)	(0.00)		(0.00)	(0.00)		(0.00)
Market Share	0.038	0.023	0.030	0.027	0.013	0.027	-0.009	-0.013	-0.005
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.03)	(0.03)	(0.03)
HHI	-0.004	0.001	-0.002	0.001	0.000	-0.002	0.017	0.008	0.007
	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)
PARTS	-0.015**	-0.019	-0.017	-0.005	0.001	0.006	-0.006	0.018	0.021
	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)
Industry dummies	YES	YES	YES	YES	YES	YES	YES	YES	YES
Country dummies	YES	YES	YES	YES	YES	YES	YES	YES	YES
Time dummies	YES	YES	YES	YES	YES	YES	YES	YES	YES
Adjusted-R^2	0.039	0.036	0.038	0.037	0.029	0.036	0.036	0.033	0.037
Number of obs.	7105	7105	7105	5879	5879	5879	4498	4498	4498

† The dependent variable is the profit margin. Standard errors in parenthesis are bootstrapped and clustered at the four-digit industry level. In specifications (OLS), the variable PARTS is a simple dummy equal to 1 if the firm participates in EU-FP (Pooled OLS is used). The variable PARTS for columns (TS1) and (TS2) corresponds to the predicted values of the first stage logit estimations (1) and (2) in table 1.8 respectively.
* Significant at the 10% level.
** Significant at the 5% level.
*** Significant at the 1% level.

TABLE 1.11: Second stage estimation results, long-term effects: Labor Productivity†

Sample	RANDOM			IC-REP			SALES-REP		
	(OLS)	(TS1)	(TS2)	(OLS)	(TS1)	(TS2)	(OLS)	(TS1)	(TS2)
δ_0	0.181***	0.370***	0.317***	0.161***	0.353***	0.337***	0.092**	0.276**	0.247*
	(0.04)	(0.12)	(0.12)	(0.03)	(0.11)	(0.11)	(0.05)	(0.14)	(0.14)
δ_1	0.138***	0.289**	0.243*	0.115***	0.347***	0.325***	0.055	0.291*	0.268*
	(0.04)	(0.14)	(0.14)	(0.04)	(0.12)	(0.12)	(0.05)	(0.16)	(0.16)
δ_2	0.146***	0.337**	0.293**	0.115***	0.371***	0.354***	0.071	0.327**	0.310*
	(0.04)	(0.14)	(0.13)	(0.04)	(0.11)	(0.11)	(0.05)	(0.16)	(0.16)
δ_3	0.134***	0.329**	0.287*	0.111***	0.370***	0.358***	0.075	0.359**	0.348**
	(0.05)	(0.15)	(0.16)	(0.04)	(0.12)	(0.12)	(0.05)	(0.17)	(0.17)
δ_4	0.111**	0.347**	0.308*	0.086*	0.378***	0.368**	0.050	0.396**	0.385**
	(0.05)	(0.16)	(0.17)	(0.05)	(0.15)	(0.15)	(0.05)	(0.16)	(0.17)

† The table presents the estimated coefficients in the regression $y_{it+\tau} = x'_{it}\beta + \delta_\tau \text{PARTS}_{it} + \varepsilon_{it+\tau}$, with $\tau = 0, \dots, 4$. Standard errors in parenthesis are bootstrapped and clustered at the four-digit industry level. In specifications (OLS), the variable PARTS is a simple dummy equal to 1 if the firm participates in EU-FP (Pooled OLS is used). The variable PARTS for columns (TS1) and (TS2) corresponds to the predicted values of the first stage logit estimations (1) and (2) in table 1.8 respectively.

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

TABLE 1.12: Second stage estimation results, long-term effects: Profit Margin†

Sample	RANDOM			IC-REP			SALES-REP		
	(OLS)	(TS1)	(TS2)	(OLS)	(TS1)	(TS2)	(OLS)	(TS1)	(TS2)
δ_0	-0.015**	-0.019	-0.017	-0.005	0.001	0.006	-0.006	0.018	0.021
	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)
δ_1	-0.013*	-0.016	-0.013	-0.003	0.011	0.014	-0.008	0.021	0.022
	(0.01)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)
δ_2	-0.011	0.002	0.004	-0.001	0.027	0.030	-0.005	0.032*	0.033*
	(0.01)	(0.02)	(0.02)	(0.01)	(0.03)	(0.03)	(0.01)	(0.02)	(0.02)
δ_3	-0.006	0.003	0.004	0.000	0.034	0.038	-0.004	0.040*	0.040*
	(0.01)	(0.02)	(0.02)	(0.01)	(0.02)	(0.03)	(0.01)	(0.02)	(0.02)
δ_4	-0.004	-0.001	-0.000	0.005	0.037	0.042	0.002	0.051**	0.053**
	(0.01)	(0.02)	(0.02)	(0.01)	(0.03)	(0.03)	(0.01)	(0.03)	(0.03)

† The table presents the estimated coefficients in the regression $y_{it+\tau} = x'_{it}\beta + \delta_\tau \text{PARTS}_{it} + \varepsilon_{it+\tau}$, with $\tau = 0, \ldots, 4$. Standard errors in parenthesis are bootstrapped and clustered at the four-digit industry level. In specifications (OLS), the variable PARTS is a simple dummy equal to 1 if the firm participates in EU-FP (Pooled OLS is used). The variable PARTS for columns (TS1) and (TS2) corresponds to the predicted values of the first stage logit estimations (1) and (2) in table 1.8 respectively.

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

Table 1.13: Analysis by RJV Size: First stage estimation results (Multinomial logit)†

Sample	Random		IC-Rep		Sales-Rep	
(1)	Large	Small	Large	Small	Large	Small
Constant	-8.889***	-4.166**	-9.593***	-4.989***	-7.020***	-3.768***
	(2.09)	(1.70)	(1.39)	(1.46)	(1.45)	(1.36)
log(Employees)	0.365***	0.452***	0.465***	0.508***	0.017	-0.040
	(0.10)	(0.11)	(0.11)	(0.11)	(0.08)	(0.09)
log(Fixed Assets Intensity)	0.257***	0.110	0.160**	0.030	0.123	-0.071
	(0.07)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)
log(Available Funding)	0.799***	0.367***	0.668***	0.271***	0.839***	0.385***
	(0.09)	(0.06)	(0.09)	(0.04)	(0.10)	(0.05)
Market Share	10.670***	3.775*	8.405***	2.947*	6.075***	1.861
	(2.49)	(2.14)	(1.86)	(1.59)	(2.03)	(2.20)
HHI	1.840	1.879*	0.751	0.917	4.072***	4.378***
	(1.38)	(1.11)	(0.99)	(1.08)	(1.36)	(1.17)
χ^2 test on						
log(Available Funding)	76.08	42.32	51.15	38.24	66.44	51.46
Pseudo-R^2	0.381		0.261		0.326	
Number of Obs.	1667		1550		1362	
(2)	Large	Small	Large	Small	Large	Small
Constant	-9.393***	-5.528***	-9.380***	-5.639***	-7.496***	-3.148
	(1.70)	(1.94)	(1.67)	(1.87)	(1.58)	(1.92)
log(Employees)	0.386***	0.475***	0.467***	0.507***	0.017	-0.036
	(0.11)	(0.11)	(0.11)	(0.11)	(0.08)	(0.08)
log(Fixed Assets Intensity)	0.204**	0.030	0.143*	-0.001	0.124	-0.114
	(0.09)	(0.09)	(0.08)	(0.09)	(0.09)	(0.09)
log(Intang Assets Intensity)	0.093	0.140**	0.037	0.060	0.003	0.072
	(0.07)	(0.06)	(0.06)	(0.06)	(0.07)	(0.06)
log(Available Funding)	0.789***	0.356***	0.665***	0.265***	0.845***	0.381***
	(0.09)	(0.06)	(0.09)	(0.04)	(0.10)	(0.05)
Market Share	10.571***	3.586	8.239***	2.737*	6.140***	1.606
	(2.58)	(2.28)	(1.90)	(1.62)	(2.04)	(2.31)
HHI	1.792	1.887*	0.725	0.847	4.016***	4.380***
	(1.38)	(1.13)	(0.98)	(1.08)	(1.36)	(1.19)
χ^2 test on						
log(Available Funding)	74.39	41.46	51.13	36.07	65.35	50.53
Pseudo-R^2	0.387		0.262		0.327	
Number of Obs.	1667		1550		1362	

† Standard errors in parenthesis and clustered at the four-digit industry level. All specifications include time, industry and country dummies. The reference outcome is not participating in the EU-FP5 IST programme.

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

Table 1.14: Analysis by RJV size, second stage estimation results: Labor Productivity†

Sample	Random			IC-Rep			Sales-Rep		
	(OLS)	(TS1)	(TS2)	(OLS)	(TS1)	(TS2)	(OLS)	(TS1)	(TS2)
Constant	1.671***	1.643***	1.704***	2.384***	2.574***	2.591***	2.181***	2.083***	1.867***
	(0.15)	(0.16)	(0.16)	(0.34)	(0.29)	(0.29)	(0.24)	(0.24)	(0.26)
log(Employees)	-0.154***	-0.172***	-0.169***	-0.118***	-0.140***	-0.140***	-0.100***	-0.101***	-0.102***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
log(Fixed Assets Intensity)	0.272***	0.277***	0.269***	0.258***	0.254***	0.251***	0.279***	0.287***	0.278***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
log(Intang Assets Intensity)	0.018**		0.012*	0.005		0.004	0.012		0.012
	(0.01)		(0.01)	(0.01)		(0.01)	(0.01)		(0.01)
Market Share	2.282***	2.542***	2.466***	0.935	0.917	0.899	0.760	0.838	0.791
	(0.63)	(0.69)	(0.68)	(0.79)	(0.83)	(0.83)	(0.51)	(0.53)	(0.52)
HHI	0.215	0.133	0.159	0.064	-0.070	-0.063	-0.070	-0.222	-0.201
	(0.18)	(0.19)	(0.19)	(0.14)	(0.16)	(0.16)	(0.16)	(0.19)	(0.18)
$PARTS_{Large}$	0.256***	0.379**	0.352**	0.507***	0.426***	0.418***	0.161***	0.335**	0.336**
	(0.05)	(0.16)	(0.16)	(0.16)	(0.14)	(0.14)	(0.05)	(0.16)	(0.15)
$PARTS_{Small}$	0.146***	0.461**	0.376**	0.322**	0.118	0.095	0.022	0.332	0.213
	(0.05)	(0.20)	(0.19)	(0.14)	(0.20)	(0.20)	(0.05)	(0.24)	(0.23)
Industry dummies	YES	YES	YES	YES	YES	YES	YES	YES	YES
Country dummies	YES	YES	YES	YES	YES	YES	YES	YES	YES
Time dummies	YES	YES	YES	YES	YES	YES	YES	YES	YES
Adjusted-R^2	0.699	0.697	0.698	0.494	0.495	0.495	0.559	0.559	0.559
Number of obs.	7105	7105	7105	5898	5898	5898	4498	4498	4498

† The dependent variable is the logarithm of labor productivity. Standard errors in parenthesis are bootstrapped and clustered at the four-digit industry level. In specifications (OLS), the variable $PART_r$ is a simple dummy equal to 1 if the firm participates in a RJV of size c (Pooled OLS is used). The variable $PART_c$ for columns (TS1) and (TS2) corresponds to the predicted values of the first stage multilogit estimations (1) and (2) in table 1.13 respectively.

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

TABLE 1.15: Analysis by RJV size, second stage estimation results: Profit Margin†

Sample	RANDOM			IC-REP			SALES-REP		
	(OLS)	(TS1)	(TS2)	(OLS)	(TS1)	(TS2)	(OLS)	(TS1)	(TS2)
Constant	-0.013	-0.001	-0.011	0.077**	0.054*	0.032	-0.025	0.001	-0.001
	(0.02)	(0.02)	(0.02)	(0.04)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)
log(Employees)	0.003	0.004**	0.004**	0.002	0.002	0.002	0.005**	0.004**	0.005**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
log(Fixed Assets Intensity)	0.009***	0.007***	0.009***	0.006**	0.003	0.006**	0.006***	0.003*	0.005**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
log(Intang Assets Intensity)	-0.002**		-0.002*	-0.004***		-0.004***	-0.002**		-0.003**
	(0.00)		(0.00)	(0.00)		(0.00)	(0.00)		(0.00)
Market Share	0.040	0.003	0.012	0.029	0.004	0.019	-0.008	-0.015	-0.009
	(0.05)	(0.05)	(0.06)	(0.05)	(0.05)	(0.05)	(0.03)	(0.03)	(0.04)
HHI	-0.005	-0.000	-0.003	0.004	0.001	-0.001	0.023	0.017	0.017
	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
$PARTS_{Large}$	-0.018*	-0.008	-0.007	0.095***	0.015	0.018	-0.008	0.020	0.025
	(0.01)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.01)	(0.02)	(0.02)
$PARTS_{Small}$	-0.012	-0.048**	-0.042*	0.083***	-0.014	-0.006	-0.003	-0.007	-0.012
	(0.01)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.01)	(0.03)	(0.03)
Industry dummies	YES	YES	YES	YES	YES	YES	YES	YES	YES
Country dummies	YES	YES	YES	YES	YES	YES	YES	YES	YES
Time dummies	YES	YES	YES	YES	YES	YES	YES	YES	YES
Adjusted-R^2	0.034	0.031	0.033	0.036	0.025	0.032	0.031	0.029	0.032
Number of obs.	7105	7105	7105	5898	5898	5898	4498	4498	4498

† The dependent variable is the profit margin. Standard errors in parenthesis are bootstrapped and clustered at the four-digit industry level. In specifications (OLS), the variable $PARTS_c$ is a simple dummy equal to 1 if the firm participates in a RJV of size c (Pooled OLS is used). The variable $PARTS_c$ for columns (TS1) and (TS2) corresponds to the predicted values of the first stage multilogit estimations (1) and (2) in table 1.13 respectively.

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

TABLE 1.16: Analysis by RJV size, second stage estimation results, long-term effects: Labor Productivity†

Sample	RANDOM			IC-REP			SALES-REP		
	(OLS)	(TS1)	(TS2)	(OLS)	(TS1)	(TS2)	(OLS)	(TS1)	(TS2)
δ_0^L	0.256***	0.379**	0.352**	0.507***	0.426***	0.418***	0.161***	0.335**	0.336**
	(0.05)	(0.16)	(0.16)	(0.16)	(0.14)	(0.14)	(0.05)	(0.16)	(0.15)
δ_0^S	0.146***	0.461**	0.376**	0.322**	0.118	0.095	0.022	0.332	0.213
	(0.05)	(0.20)	(0.19)	(0.14)	(0.20)	(0.20)	(0.05)	(0.24)	(0.23)
δ_1^L	0.224***	0.313*	0.293*	0.417***	0.474***	0.464***	0.130***	0.250	0.263
	(0.05)	(0.17)	(0.16)	(0.16)	(0.15)	(0.15)	(0.05)	(0.19)	(0.18)
δ_1^S	0.095*	0.380*	0.301	0.217	0.133	0.097	-0.015	0.441*	0.329
	(0.05)	(0.21)	(0.21)	(0.14)	(0.25)	(0.25)	(0.06)	(0.27)	(0.25)
δ_2^L	0.228***	0.370**	0.349**	0.453***	0.566***	0.559***	0.146***	0.300	0.309
	(0.06)	(0.17)	(0.17)	(0.17)	(0.17)	(0.17)	(0.05)	(0.20)	(0.20)
δ_2^S	0.101*	0.372*	0.306	0.246	0.115	0.081	0.004	0.384	0.302
	(0.05)	(0.22)	(0.21)	(0.15)	(0.29)	(0.29)	(0.05)	(0.26)	(0.24)
δ_3^L	0.229***	0.385**	0.371**	0.453***	0.598***	0.596***	0.155***	0.358*	0.371*
	(0.06)	(0.19)	(0.18)	(0.17)	(0.16)	(0.16)	(0.06)	(0.21)	(0.20)
δ_3^S	0.084	0.336	0.252	0.253	0.098	0.059	0.004	0.265	0.187
	(0.06)	(0.25)	(0.25)	(0.16)	(0.34)	(0.34)	(0.06)	(0.32)	(0.30)
δ_4^L	0.180**	0.409**	0.388*	0.430**	0.617***	0.618***	0.096	0.436**	0.441**
	(0.07)	(0.21)	(0.20)	(0.20)	(0.19)	(0.19)	(0.06)	(0.21)	(0.21)
δ_4^S	0.068	0.320	0.264	0.290	0.298	0.264	-0.005	0.183	0.116
	(0.06)	(0.31)	(0.31)	(0.19)	(0.42)	(0.42)	(0.06)	(0.34)	(0.32)

† The table presents the estimated coefficients in the regression $y_{it+\tau} = x'_{it}\beta + \delta_\tau^L \cdot \text{PARTS}_{Lit} + \delta_\tau^S \cdot \text{PARTS}_{Sit} + \varepsilon_{it+\tau}$, with $\tau = 0, \ldots, 4$. Standard errors in parenthesis are bootstrapped and clustered at the four-digit industry level. In specifications (OLS), the variable PARTS_c is a simple dummy equal to 1 if the firm participates in a RJV of size c (Pooled OLS is used). The variable PARTS_c for columns (TS1) and (TS2) corresponds to the predicted values of the first stage multilogit estimations (1) and (2) in table 1.13 respectively.

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

TABLE 1.17: Analysis by RJV size, second stage estimation results, long-term effects: Profit Margin†

Sample	RANDOM			IC-REP			SALES-REP		
	(OLS)	(TS1)	(TS2)	(OLS)	(TS1)	(TS2)	(OLS)	(TS1)	(TS2)
δ_0^L	-0.018*	-0.008	-0.007	0.095***	0.015	0.018	-0.008	0.020	0.025
	(0.01)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.01)	(0.02)	(0.02)
δ_0^S	-0.012	-0.048**	-0.042*	0.083***	-0.014	-0.006	-0.003	-0.007	-0.012
	(0.01)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.01)	(0.03)	(0.03)
δ_1^L	-0.017	-0.001	0.001	0.096***	0.030	0.031	-0.011	0.025	0.029
	(0.01)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.01)	(0.02)	(0.02)
δ_1^S	-0.010	-0.049*	-0.042*	0.086***	-0.006	-0.001	-0.005	-0.001	-0.009
	(0.01)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.01)	(0.04)	(0.04)
δ_2^L	-0.014	0.021	0.023	0.078**	0.062	0.064	-0.006	0.045	0.049*
	(0.01)	(0.02)	(0.02)	(0.04)	(0.04)	(0.04)	(0.01)	(0.03)	(0.03)
δ_2^S	-0.009	-0.050*	-0.045*	0.067**	-0.019	-0.014	-0.003	-0.015	-0.022
	(0.01)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.01)	(0.04)	(0.04)
δ_3^L	-0.007	0.027	0.030	0.073**	0.070*	0.072*	-0.005	0.045*	0.048*
	(0.01)	(0.02)	(0.02)	(0.04)	(0.04)	(0.04)	(0.01)	(0.03)	(0.03)
δ_3^S	-0.006	-0.055*	-0.054*	0.062**	-0.017	-0.013	-0.001	-0.005	-0.013
	(0.01)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.01)	(0.05)	(0.05)
δ_4^L	-0.009	0.032	0.034	0.089**	0.082**	0.083**	-0.002	0.062**	0.065**
	(0.01)	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)	(0.01)	(0.03)	(0.03)
δ_4^S	0.001	-0.060**	-0.058*	0.084**	-0.030	-0.025	0.008	-0.014	-0.019
	(0.01)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.01)	(0.06)	(0.06)

† The table presents the estimated coefficients in the regression $y_{it+\tau} = x'_{it}\beta + \delta_\tau^L \cdot \text{PARTS}_{Lit} + \delta_\tau^S \cdot \text{PARTS}_{Sit} + \varepsilon_{it+\tau}$, with $\tau = 0, \ldots, 4$. Standard errors in parenthesis are bootstrapped and clustered at the four-digit industry level. In specifications (OLS), the variable PARTS_c is a simple dummy equal to 1 if the firm participates in a RJV of size c (Pooled OLS is used). The variable PARTS_c for columns (TS1) and (TS2) corresponds to the predicted values of the first stage multilogit estimations (1) and (2) in table 1.13 respectively.

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

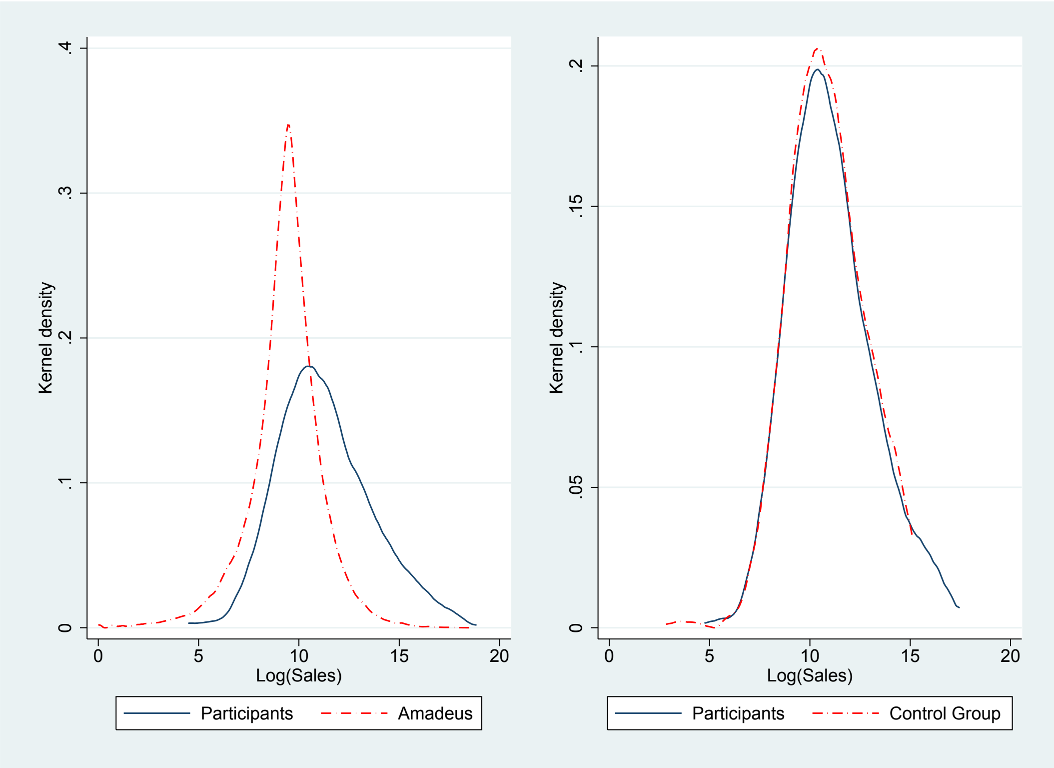

FIGURE 1.1: Sales distributions for participants, Amadeus and Control Groups in 1999.

Chapter 2

Knowledge Spillovers in Cost-Reduction Incentives

2.1 Introduction

This goal of this paper is to identify and measure the relevance of knowledge spillovers in the French urban transport industry. In each French city of significant size, the local authority regulates and monitors the activity of a single operator that provides the transport services on the urban network within a regulatory framework. The latter takes the form of a written contract that defines the payment and cost-reimbursement rules between the parties. Two types of regulatory contracts are observed in practice, namely fixed-price and cost-plus schemes. Under fixed-price contracts, the operator receives subsidies to cover ex-ante (expected) operating deficits, while under cost-plus regulation, subsidies are provided by the local authorities to finance ex-post (realized) deficits. As it is well known, each type of regulatory rule has an impact on operating costs since it entails different levels of incentives in terms of effort in cost reduction activities. In particular, fixed-price contracts provide powerful incentives to reduce operating costs.

The regulator does not observe the technological efficiency or the cost reduction activity of the operator. In France, local authorities have been historically blamed for their laxness in assessing operating costs, mainly because of their lack of knowledge and experience of transportation economics and technologies, and/or because of their limited capacity of monitoring and auditing complex operating activities. These considerations prevent them to adequately assess the effort of operators in providing appropriate and competent solutions to cost and network inefficiencies.

A distinguishing feature of the French urban transport industry is that about eighty percent of the operators are owned by three large companies. The transport services provided in different networks by operators belonging to the same industrial group are therefore, in essence, provided by the same firm. This peculiarity suggests that these companies may benefit from spillovers by operating several networks in different localities. In other words, the economic activity involved in one specific network may affect the economic activity in other networks operated by the same company. In the specific context of the urban transport industry, we expect these spillovers or externalities to take place when a cost-reducing activity developed by one of the operators reaches other parts of the group. As a consequence of this positive externality, all networks operated by the firms belonging to the group benefit from the cost-reducing efforts provided by any of the operators. In the French urban transportation industry, operators have engineer teams in each network that are responsible for the improvement of the operator's productivity. When operators belong to the same group, the new methods and procedures that they develop can potentially be used by the entire company, therefore improving productivity and/or reducing operating costs. To reflect the fact that these spillovers could be related to a large array of know-how generated by the firm, ranging from technological to organizational, we will refer to them as knowledge spillovers throughout the text.

This paper is aimed at identifying and measuring the importance of these knowledge spillovers in the French urban transport industry and their impact on the efficiency of operators.

Our work shares features with different strands of the empirical literature on both regulation and firm performance. First, our paper belongs to the recent empirical literature on incentive regulatory policies. Considering contract in the transportation industry, Gagnepain and Ivaldi (2002) focus on the incentives effects of cost-plus and fixed-price contracts on the cost-reducing activities of operators and compare the welfare level of observed regulatory practices with the one that would have been obtained using first-best and second-best regulatory policies. Gagnepain and Ivaldi (2010) build on a similar model and incorporate political aspects of regulation to analyze the determinants of contract choices. Gagnepain et al. (2013) analyze the dynamics aspects of regulation by assessing the cost of renegotiation in the French transportation industry. Our models builds on this previous literature by focusing on the incentives of cost-plus and fixed-price contracts and extends it by modeling the knowledge spillovers that result from these incentives when transport operators are linked to each other. From that perspective, our paper is one of the first to take into account knowledge spillovers in a regulation context.

Dealing with technology and knowledge externalities, our paper also relates to the empirical literature on R&D knowledge spillovers. However, most of the latter deals with knowledge spillovers across firms rather than within firms (Bernstein and Nadiri, 1989; Griliches, 1992;

Bloom et al., 2013). Most related to our paper is the work of Klette (1996). Using data on Norwegian manufacturing firms, he analyzes the relationship between firm performance and R&D. In addition to being able to identify R&D at the line-of-business level within each firm, he can identify firms that belong to the same "interlocking group of firms".[1] His results show significant spillover effects across different lines of business (e.g. chemicals or metal products) within a firm but also reveal significant spillovers for activities within a line of business that are carried out by different firms within the same group.

Our work also relates to part of the literature on knowledge transfer within the firm. In a seminal paper, Szulanski (1996) analyzes firms' ability to transfer best practices internally. Using survey data on best practice transfers in eight different companies, he shows the major barriers to internal knowledge transfer to be knowledge-related factors such as the recipient lack of absorptive capacity. Although we are not able to explicitly measure the knowledge transfers within groups in our analysis, we can still account for some form of network-specific absorptive capacity in our analysis and find evidence of positive knowledge spillovers across networks linked to the same group.[2] Another related paper is Darr et al. (1995), which analyze knowledge transfer acquired through learning-by-doing in service organization. Focusing on 36 pizza stores owned by 10 different franchisees, they find evidence of knowledge transfer across stores owned by the same franchisee but not across stores owned by different franchisees. Related to this, some papers have also highlighted the importance of free-riding among the different franchisees of a given chain (Brickley, 1999; Lafontaine and Slade, 1997, 2007). Indeed, a franchisee has incentives to free-ride on the tradename of the franchisor given that her effort is private while the benefits will accrue to all the members of the chain. This closely relates to our case where each local network privately pays the cost of its effort which will benefit (at least partially) all members of the group.

We build and estimate a structural cost regulation model with asymmetric information that includes knowledge spillovers (i.e. externalities) of operating different networks. Our goal is to identify the latter, and see how they influence firms' decisions of exerting effort to reduce their operating costs. Our model provides us with estimates of the firms' inefficiencies, the effort of the managers and the technology.

In each given city, there's a single network operator in charge of providing the transportation service. This operator can be part of one of the largest industrial groups or it can be independent. In either case, each operator has a local manager who takes care of running the network and decides on the effort to be exerted to reduce the operating costs of the local transportation activity. In our model, each operator in each given city faces different

[1] According to the author, an interlocking group of firm is characterized by a parent company and all subsidiaries in which the parent company owns a majority share of equity.

[2] In particular, we are able to construct indices that relate to the structural differences between a given network and the remaining networks from the same group. They are presented in greater detail below.

incentives to reduce these costs. First, the type of regulatory contract (fixed-price vs cost-plus) is a crucial determinant of cost reducing activities. The second class of incentive is the one driven by the knowledge spillovers that follow from operators being part of a same group, if any. The econometric task consists then in recovering the parameters of a static model of cost regulation under incomplete information, and testing for the relevance of knowledge spillovers.

Our results show that knowledge spillovers are indeed relevant in the French urban transport industry. Note that we do not model the regulation decision by the authority, i.e. we take the regulatory mechanisms as exogenously given to estimate the model's parameters.[3] Likewise, we leave dynamic considerations out of our framework. In particular, we do not address the ability of the regulator to commit not to use the information on the operator's cost from one regulatory period to another.[4] Here, our aim is to focus on the knowledge spillovers that arise from the operators' group structure within a static framework.

The organization of the paper is as follows: Section 2 describes the regulation of urban transportation in France in more detail. Section 3 presents the contracts that are implemented during our period of observation and describes the structure of the different existing industrial groups in the French urban transport industry. Section 4 discusses the assumptions that are maintained throughout the paper. Section 5 presents our model of cost regulation which encompasses the main features of urban transportation and the environment in which network operators make their decisions. Section 6 then develops a formal specification of the cost function to be estimated. Section 7 is devoted to the construction of the variables and the presentation of our results. Section 8 evaluates the cost gains of adding operators to a group and further simulates the cost gains following a merger between two groups. Section 9 provides a summary and some concluding remarks.

2.2 The French Urban Transportation Industry

As in most countries around the world, urban transportation in France is a regulated activity. Local transportation networks cover each urban area of significant size, and for each network, a local authority (a municipality, a group of municipalities or a district) is in charge to regulate an operator which has been selected to provide the transportation service. Regulatory rules prevent the presence of several suppliers of transportation services on the same urban network, and each network is therefore operated by a single operator.

[3]See Gagnepain and Ivaldi (2010) for a political model of regulation that incorporates the determinants of contract choices.

[4]These issues are addressed in Gagnepain et al. (2013).

The 1982 Transportation Law was enacted to facilitate decentralized decision-making on urban transportation and to provide guidelines for regulation. As a result, each local authority now organizes its own transportation system by setting route and fare structures, capacity, quality of service, conditions for subsidizing the service, levels of investment and ownership nature. The local authority may decide to operate the network directly or to require the services of a transport service provider. In the latter case, a formal contract defines the regulatory rules that the operator must follow as well as the payment and/or cost-reimbursement scheme between the authority (the principal) and the operator (the agent).

In most urban areas, operating costs are on average twice as high as commercial revenues. Budgets are therefore rarely balanced without subsidies. One reason is that operators face universal service obligations and must operate in low demand areas. Low prices are maintained to ensure affordable access to all consumers of public transportation. Moreover, special fares are given to targeted groups like seniors and students. Subsidies come from the State's budget, the local authority's budget, and a special tax paid by local firms (employing more than nine full-time workers). They are not necessarily paid directly to the operator. In addition to the price distortions causing deficits, informational asymmetries that affect the cost side and lead to inefficiencies make it more difficult to resume these deficits. We return on these points more in detail below.

A distinguishing feature of France compared to most other OECD countries is that about eighty percent of local operators are private and are owned by three large companies, two of them being private while the third one is semi-public.[5] In 2002, these companies, with their respective ownership structures and market shares (in terms of number of networks operated) were Keolis (private, 30%), Transdev (semi-public, 19%), Connex (private, 25%). In addition there are a small private group, Agir, and a few public firms controlled by local governments. The next section is dedicated to a more detailed presentation of the structure of these groups.

Our objective is to take these features of the urban transport industry into account and to perform an analysis of the observed regulatory schemes within a principal-agent setting. This requires a database that provides information on both the performance and the organization of the French urban transport industry. Such a database was created in the early 1980s from an annual survey conducted by the *Centre d'Etude et de Recherche du Transport Urbain* (CERTU, Lyon) with the support of the *Groupement des Autorités Responsables du Transport* (GART, Paris), a nationwide trade organization that gathers most of the local authorities in charge of a urban transport network. In France, this rich source is a unique tool for comparing observed regulatory schemes both across year and over time. In our econometric

[5]For an overview of the regulation of urban transportation systems in the different countries of the European Union, the United States and Japan, see IDEI (1999).

analysis, we consider the regulatory scheme adopted in each urban area during a year as a realization of the same regulatory contract. The sample does not include the largest networks of France, i.e., Paris, Lyon and Marseille, as they are not covered by the survey. Overall, the panel data set covers 87 different urban transport networks over the period 1987-2001.

We now turn to a more detailed description of the contractual relationships and of the urban transport group structures.

2.3 Regulatory Contracts and Transport Groups

Two types of regulatory contracts are implemented in the French urban transport industry, namely fixed-price and cost-plus schemes. Over the period of observation, fixed-price contracts are employed in 61% of the cases, as shown in table 2.1. Under fixed-price contracts, operators receive subsidies to finance the expected operating deficits, while under cost-plus schemes, subsidies are paid to the local authorities to finance ex-post deficits. Hence fixed-price regimes are very high-powered incentive schemes, while cost-plus regimes do not provide any incentives for cost reduction. For the same network, the regulatory scheme may switch from cost-plus to fixed-price or from fixed-price to cost-plus between two contract periods. We indeed count twenty-three changes of regulatory regimes, eighteen of them being switches from cost-plus to fixed-price regimes. These changes may occur either during the term of a local authority which decides for a shift in the regulatory framework, or after the election of a new local government.

As already mentioned, and as shown in table 2.1, about eighty percent of local transport service operators are owned by three larger industrial groups; Keolis, Connex and Transdev. Industrial groups of urban transport have a long history of mergers in France. Keolis was born out of the merger of several companies created in the beginning of the 20th century. The *Société des transports automobiles*, created in 1908, its subsidiary (the *Société générale des transports départementeaux*) and the company *Lesexel*, founded in 1911 to help on the development of tramways, merged to form the VIA-GTI company, mainly focused on urban transport. In the meantime, another company, *Cariane,* was specialized in the French interurban transport. Ultimately, VIA-GTI and *Cariane* merged in 2001 to give birth to Keolis.[6] The industrial group Connex was born out of the merger of the *Compagnie Générale Française des Transports et Entreprises* (CGFTE) and the *Compagnie Générale d'Entreprises Automobiles* (CGEA) in the late 1980's.[7] The company was ultimately renamed *Veolia Transport* in 2005. Finally, the Transdev group was created in 1955. On March 3rd 2011, it merged with *Veolia Transport* to give birth to *Veolia Transdev.*

[6] http://www.keolis.com/en/about-us/key-facts/history.html

[7] The company actually decided to take on the name Connex in 2000. For more details, see http://www.connex.info/tmpl/ExtensionPage____2778.aspx?epslanguage=ML.

Table 2.1 presents figures on the distribution of the operators in our sample according to their affiliation to one of the three groups and to the type of contract they face. While 61% of all the networks are under a FP regime, this figure amounts to 64.5% when focusing on the operators belonging to a group and to 48% for independent networks. Although it seems that belonging to a group is an important determinant of being regulated under a FP contract, a closer look at the frequency of FP contracts within groups indicates that this figure comes mainly from the Transdev group, where more than 90% of the networks it operates are under a FP scheme. The overall proportion of networks operated by firms belonging to one of the three major groups is of 78.6%. For networks regulated under a FP regime, 83.1% are operated by firms belonging to a group. For networks under CP contracts, this figure amounts to 71.6%.

For each urban transport network, the automatic renewal of the contract between the local authority and the operator in place was effectively ended, by law, in 1993. Since then, local authorities are required to use beauty contests to allocate the construction and management of infrastructures of urban transportation. In practice, however, very few networks have experienced changes of operators from one regulatory period to another. Over the period covered by our analysis, only 5 networks have decided to get rid of their operators to select another company. Out of these, two changed from being operated by a firm belonging to a group to a being operated by an independent firm, while only one network changed from being operated by an independent firm to being operated by a firm belonging to a group. Finally, only 2 networks saw their operator change from a firm belonging to a given group to a firm belonging to another group.

As a matter of fact, the different operators mostly avoided head to head competition and generally put tenders for markets in distinct urban areas. By committing to distinct geographical areas, the three main groups succeeded in reducing the degree of competition in the awarding of transport operations in urban areas where the regulatory contract comes to an end. Competitive tendering is therefore not a relevant issue in this sector, and ex-ante competition is not so fierce. Finally, these groups also operate other municipal services such as water distribution or garbage collection, which makes it even harder for public authorities to credibly punish operators following bad performances in the provision of transport services. It follows that group structures are rather stable both across networks and over time in our sample.

These urban transport industry features constitute the core of our analysis on the knowledge spillovers among operators belonging to the same group and inspire the construction of the structural model of regulation that we present below. Before going to the construction of our economic model, we introduce some assumptions that we now present in detail.

2.4 Delineating the Scope of the Study

The organization and structure of the urban transportation industry in France as described above motivates the following assumptions.

Assumption 1: The network operator has private information about its technology, and the authority does not observe its effort to improve productivity or to reduce costs.

We assume that the network operator has private information about its innate technology (adverse selection) and that its cost-reducing effort is non-observable (moral hazard). Because French local authorities exercise their new powers on transportation policy since the enactment of the 1982 Law only, and since they usually face serious financial difficulties, their limited auditing capacities is recognized among practitioners. A powerful and well-performed audit system needs effort, time and money. French experts on urban transportation blame local authorities for being too lax in assessing operating costs, mainly because of a lack of knowledge of the technology.[8] The number of buses required for a specific network, the costs incurred on each route, the fuel consumption of buses (which is highly dependent on drivers' skills), the drivers' behavior toward customers, the effect of traffic congestion on costs, are all aspects for which operators have much more data and a better understanding than public authorities. This suggests the presence of adverse selection on innate technology in the first place. Given the technical complexity of these issues, it should be even harder for the local authority to assess whether and to what extent operators undertake efforts to provide appropriate and efficient management. Moral hazard arises quite naturally on top of the adverse selection problem. When compounded, those informational asymmetries play a crucial role in the design of contractual arrangements and financial objectives.[9]

Assumption 2: Regulatory schemes and operators' efficiency levels are exogenous

According to the new theory of regulation, when contractual relationships are characterized by informational asymmetries, a welfare-maximizing regulator applies the revelation principle to provide the operator with incentives to reveal its true efficiency level. This mechanism

[8] The French urban transport expert O. Domenach has argued that *"the regulator is not able of determining the number of buses which is necessary to run the network. The same comment can be made regarding the fuel consumption of each bus. The regulators are generally general practitioners instead of transport professionals. Hence, the (re)negotiation of contracts between regulators and operators is not fair.* See Domenach (1987). A more recent report on the weak capabilities of expertise of the local governments and the lack of ex ante competition in the industry is proposed by the French court of auditors de Comptes (2005); for more details see http://www.ccomptes.fr/content/download/2454/24573/version/2/file/RapportTransports.pdf.

[9] Gagnepain and Ivaldi (2002) confirmed through a test that adverse selection and moral hazard are two important features of the industry. They showed that a regulatory framework which encompasses these two ingredients performs well to explain the data.

can be decentralized through a menu of linear contracts and avoids leaving excessive rents. Each operator facing such a menu chooses the contract that corresponds to his own efficiency level. In this context, the most efficient firm chooses the highest-powered incentive scheme, i.e. a fixed-price contract, while the most inefficient firm chooses the lowest-powered incentive scheme, i.e. a cost-plus contract. Between these two extremes are incentive schemes chosen by firms with intermediate efficiency levels (Laffont and Tirole, 1993).

Does this framework apply to the French urban transport industry? If it did, fixed-price and cost-plus contracts would be extreme cases of a menu and would be chosen by the most efficient and the most inefficient firms, respectively. Since current rules apply to any companies (even the ones with intermediate efficiency levels) and since the real world cannot be confined to fully efficient or inefficient firms, one must conclude *a priori* that observed contracts do not include any revelation principle, and cost-plus and fixed-price schemes are equally proposed to operators without paying any attention to their efficiency level. It is therefore realistic to assume that regulatory schemes are not driven by the intrinsic characteristics and efficiency levels of large service companies and of network operators.

Assumption 3: An operators belonging to an industrial group benefits from the cost reducing activities of the remaining operators of the group.

In each network, the existence of inefficiencies may lead to higher operating costs than the levels defined by the cost frontier. Firms can, however, undertake cost-reducing activities to overcome these inefficiencies. They can, for instance, engage in process research and development, or managers can spend time and effort in improving the location of inputs within the network. They can as well attempt to find cheaper suppliers, bargain better procurement contracts, subcontract non-essential activities, monitor employees, or solve potential labor conflicts.

We assume that operators belonging to one of the groups presented above (i.e. Keolis, Transdev or Connex) will be affected by actions taken in other networks operated by another member of the same group. In each network, the group has a local manager (or team of managers) who takes care of running the network and has decision rights on the effort to be exerted in order to decrease operating costs. Given this decision-making configuration, we expect actions related to cost-reducing activities taken in a specific network to generate a positive externality on the operating costs of the remaining operators of the group. The main idea is that knowledge generated in a given location can be processed by the group's headquarters and later be transmitted and used in another network operated by the group. For instance, the results of process R&D obtained in one location can spill-over to another operator through the group's headquarters. This operator would therefore benefit from (part of) this R&D without investing as much effort as it would have to if it were independent. Similarly, the effort incurred to find a cheaper supplier in one network may reduce the need to look for a cheaper supplier in another city. The bargaining of procurement contracts may also be easier if the operator belongs to a group with relevant experience in other networks. Likewise, methods to efficiently monitor employees could also be learned in a given place and transmitted to another. In that sense, a network belonging to a group will benefit from positive externalities coming from the effort exerted in all the remaining networks of the group. To reflect the fact that these spillovers could be related to a large array of know-how generated by the firm, ranging from technological to organizational, we refer to them as knowledge spillovers.

We propose to estimate a structural cost function that accounts for the regulatory scheme faced by the operator as well as for the structure of the group it belongs to, if any. This allows us to test for the relevance of knowledge spillovers among operators in the French urban transport industry.[10]

We now turn to the construction of our structural model of regulation.

2.5 The Economic Model

We now present our model of regulation of the urban transport industry. Starting from the technology associated with the transportation activity, we first define the primal operating cost function, which is conditional on the cost-reducing activity of the operator. We describe how the contract types and the structure of the transport groups affect the operators' choice of cost-reducing effort. Once the optimal level of effort is determined, we plug it back into the conditional cost function to obtain the final cost function that captures all the relevant incentives affecting the activity of the firm.

Technology and primal cost function

To provide the required level of services Q, the transit firm (the operator) needs to combine variable and fixed inputs. Let $w = (w_L, w_M)$ be the price of variable inputs, namely labor (L)

[10]Three additional remarks should be made. First, private information on demand is not a relevant issue in our industry. Local governments are well informed about the transportation needs of citizens. The number of trips performed over a certain period is easily observed, and the regulator has a very precise idea of how the socio-demographic characteristics of a urban area fluctuate over time. Given the level of demand, the regulator sets the service capacity provided by the operator. Second, we do not address the issue of determining what should be the optimal rate-of-return on capital. The rolling stock is owned by the local government for a vast majority of networks. In this case, the regulator is responsible for renewing the vehicles, as well as guaranteeing a certain level of capital quality. Finally, we rule out the possibility of risk sharing in the contractual relationships between the operators and the regulators since the provision of transport services does not entail unpredictable cost fluctuations for the operators.

and materials (M). Let K and I be, respectively, the stock of capital and the infrastructure used by the operator, which are both fixed in the short run. The production process is then represented with the production function $Q = f(K, I, L, M|\lambda)$, where λ is a vector of parameters characterizing the technology in the production process. Note that both L and M are the efficient levels of inputs, which are only observable to the operator. We denote by C the observed operating cost of each firm. As the stock of capital K and the size of the infrastructure I are determined by the regulator, our cost function is determined in the short run, and is conditional on the stock of capital and on the size of the infrastructure.[11] Each operator chooses the cost-minimizing input allocation subject to technological constraints, which leads to a cost function of the following form:

$$C_i^0 = C_i^0(w_i, Q_i, I_i, K_i|\beta), \tag{2.1}$$

where β is a vector of parameters characterizing the cost function. In reality, the actual operating cost may differ from the minimum operating cost defined by (2.1). Inefficiencies may prevent operators from reaching the required level of service Q at the minimum cost, which will result in upward distorted costs. To counterbalance these inefficiencies however, firms can undertake cost-reducing activities. They can engage in process research and development, or managers can spend time and effort in improving the location of inputs within the network. They can as well attempt to find cheaper suppliers, bargain better procurement contracts, subcontract non-essential activities, monitor employees, or solve potential labor conflicts. Whatever these cost-reducing activities may be, we will refer to them as effort.

A distinguishing feature of the French urban transport system is that about 80% of local operators are private and are owned by three large industrial groups. In 2002, these companies, with their respective market shares (in terms of number of networks operated) were Keolis (30%), Transdev (19%) and Connex (25%). Hence a given firm i operating a specific network can be either independent or belong to one of these larger companies. Each of these industrial groups $g = \{Keolis, Transdev, Connex\}$ operates a set of urban networks $N_g = \{1, \ldots, n_g\}$. While production inputs are exclusively network specific, we assume the inefficiencies to affect all the n_g networks of a given group g. Likewise, we expect the cost-reducing efforts exerted in a given network to affect the operating cost of other firms belonging to the same industrial group. These knowledge spillovers are, however, not present for an independent network. We return to these points more in detail below.

Denote by θ_g the intrinsic inefficiency level of each of the n_g networks of group g, and let θ be the intrinsic inefficiency level of an independent network. We denote the effort level of

[11]In practice, the operator plays a role in the choice of investment, which, potentially, introduces another dimension that can be affected by information asymmetries. Our understanding of the industry is that this question is of second-order since, for instance, the production of new buses, which could have a drastic impact on the efficiency of the transport network, takes time and refers to periods longer than regulatory periods.

firm i belonging to group g by e_{ig}, and let e_{-ig} denote the effort of the remaining networks belonging to the same group. Let e_i be the effort level of an independent network i. Note that both the inefficiency and the effort levels are unobservable to the regulator and to the econometrician. Each operator therefore faces a cost function which provides the frontier of minimum operating costs conditional on the levels of capital, infrastructure, inefficiency, effort and group structure. Specifically, operator i faces a cost function of the form:

$$C_i(C_i^0, \theta, e|\beta) = \begin{cases} C_i^0 \times \phi\left(\theta, e_i\right) & \text{if } i \text{ is independent} \\ \\ C_i^0 \times \phi\left(\theta_g, e_{ig}, \kappa_{ig} e_{-ig}\right) & \text{if } i \in N_g, \end{cases} \tag{2.2}$$

where $\phi\left(\theta, e\right)$ is a continuous function that is increasing in θ and decreasing in e. κ_{ig} is a parameter measuring the knowledge spillovers obtained by operator i for being linked to the remaining operators belonging to group g. Notice how we define the κ_{ig} parameter to depend both on network and group characteristics, as we expect networks within a same group to benefit asymmetrically from knowledge spillovers. Note that while the inefficiency parameter θ is exogenous, the cost reducing effort is a choice variable which will depend on both the contract that the firm faces and on the structure of the group it belongs to, if any. We next turn to the operator's effort decision and to the construction of the structural cost function.

Incentives, knowledge spillovers and the optimal level of effort

Two main aspects dictate the incentives that each operator faces to reduce costs through the conditional cost function (2.2). The first environment's characteristic that affects the operator's incentives to reduce its costs comes from the regulatory pressure, defined by the type of contract that the operator faces. Two regulatory contracts are observed in practice, namely fixed-price (FP) and cost-plus (CP). Under a fixed-price contract, the operator is residual claimant for effort. It obtains an ex-ante subsidy t^{FP} equal to the expected balanced budget, which is the difference between expected costs and expected revenues. This contract is a very high-powered incentive scheme as the operator is now responsible for insufficient revenues and cost overruns. With the cost-plus contract, the public authority receives the commercial revenue $R(q)$, and receives an ex-post subsidy t^{CP} that reimburses the firm's total ex-post operational cost C. The firm is therefore not residual claimant for effort and this contract is a very low powered incentive scheme. Under this regime, firms have no incentives to produce efficiently. The operator can, under both types of contracts, exert effort e to reduce its operating cost C. The cost reduction activity induces an internal cost $\psi(e)$.

The second aspect of the economic environment that affects the incentives to reduce costs is whether the operator belongs to an industrial group or whether it is independent. Each network that belongs to a group has a local manager (or team of managers) who takes care of running the network and has decision rights on the effort to be exerted in order to decrease operating costs. Each manager is self-interested in the sense that their objective is to maximize local profits, but their cost-reducing efforts can reach other networks from the group through the group's headquarters and therefore result beneficial to them. In other words, while a group is in effect operating several networks simultaneously, we assume that decisions on cost-reducing activities are made locally and independently. There are several reasons to justify this assumption. First, network operators are not able to unilaterally decide on the type of regulatory contract they will face as this is the result of a negotiation process between the operator and the local authority. A given group could therefore never choose which type of contract to face in a given network, much less in all the different network it operates. Second, this negotiation process is always carried out locally and at different points in time, implying that contracts do not end at the same time in different networks. In other words, even if it wanted to, a group would not be able to coordinate the negotiation of contracts in the different networks it operates given this timing difference. Moreover, local authorities also change over time, meaning that the type of contract resulting from the negotiation process in a given period would not necessarily translate into the same outcome in the regulatory period.[12] These considerations imply that contract types will change across the networks of a given group at different points in time. This sequential decision configuration therefore renders a centralized planning of effort impossible and implies a local choice of effort at the network level. Finally, note also that decisions on cost-reduction activities cannot solely be based on group-level knowledge of the transportation activity, but must also depend on specific knowledge of the local network where the service is provided. For example, the number of buses required for a specific network or the drivers' skills and behavior towards customers in a specific location are all aspects for which local managers have much more information and knowledge than the group's headquarters. As such, it seems natural to assume that cost-reduction activities are first independently carried out at the local level. The results from these activities are then naturally transmitted to the group's headquarters.

Given this decision-making configuration, we expect actions related to cost-reducing activities taken in a specific network to generate a positive spillover on the operating costs of the remaining operators of the group. These spillovers are a central point in our model. We have in mind new methods, procedures or general knowledge that are generated in one network and can be transmitted to another via the group's structure. For instance, the results of process R&D obtained in one location can spill-over to another operator through

[12] See Gagnepain and Ivaldi (2010) for an analysis of how the political color of the local authority affects the negotiation of regulatory contracts in the French transportation industry.

the group's headquarters. This operator would therefore benefit from (part of) this R&D without investing as much effort as it would have to if it were independent. Similarly, the effort incurred to find a cheaper supplier in one network may reduce the need to look for a cheaper supplier in another city. The bargaining of procurement contracts may also be easier if the operator belongs to a group with relevant experience in other networks. Likewise, methods to efficiently monitor employees could also be learned in a given place and transmitted to another. Note, however, that the knowledge generated in a given location may not necessarily be perfectly transferable or applicable to another network of the group. In particular, its application and use in other locations will depend on specific network characteristics and capabilities (such as input and/or network structure).

While the necessary amount of effort to reduce inefficiencies is affected by the group structure a firm belongs to, we assume that it is still expensive to exert a given amount of effort. In other words, we assume that the marginal cost of effort is not affected by the group structure. For a given firm that belongs to group, this means that while the efforts exerted in the remaining networks will affect its operational costs, the function $\psi(e)$ will only depend on its own effort. These network effects are not present for an independent network and its effort level will therefore only depend on the type of contract it faces.

We now explicitly take into account these incentives through the cost function (2.1) that is conditional on inefficiency θ and the effort level e. We first derive the optimal level of effort for each operator and check how this effort depends on the incentives mentioned above. Second, we plug back this equilibrium level of effort into the conditional cost function. This will lead us to an unconditional structural cost function that can be estimated. Accounting for these changes in incentives through the cost structure enables us to reduce the source of misspecification and avoid biases in the estimation of the technological parameters.

Each industrial group g operates a set of urban networks N_g in quantity $\text{card}\,(N_g) = n_g$. Let N_g^{fp} denote the set of networks that the group g operates under a *FP* contract, which entails $\text{card}\left(N_g^{fp}\right) = n_g^{fp}$ networks. Similarly, let N_g^{cp} denote the set of networks that the group g operates under a *CP* contract, which entails $\text{card}\,(N_g^{cp}) = n_g^{cp}$ networks. Hence, for each group g we have that $n_g = n_g^{fp} + n_g^{cp}$.

Under a fixed-price contract, each operator i determines the optimal effort level that maximizes the objective function

$$\pi_i = \begin{cases} t_i^{fp} + R\left(q_i\right) - C_i(C_i^0, \theta, e_i|\beta) - \psi\left(e_i, \alpha\right) & \text{if } i \text{ is independent} \\ \\ t_i^{fp} + R\left(q_i\right) - C_i(C_i^0, \theta_g, e_{ig}, \kappa_{ig} e_{-ig}|\beta) - \psi\left(e_{ig}, \alpha\right) & \text{if } i \in N_g, \end{cases} \tag{2.3}$$

where $R(q) = p(q)q$ denotes revenue and q measures transport demand.[13] If the network is independent, the optimal effort level e_i^{fp} that maximizes its profit in (2.3)

is determined by the following first order condition:

$$-\frac{\partial C_i(C_i^0, \theta, e_i|\beta)}{\partial e_i} = \frac{\partial \psi_i(e_i, \alpha)}{\partial e_i}, \tag{2.4}$$

which implies that the optimal level of effort e_i^{fp} is chosen to equalize marginal cost savings with the marginal disutility of effort.

For a firm belonging to a group, the optimal effort level will depend on the effort exerted by the remaining members of the group. Each of the networks belonging to group g that are under a FP contract satisfy the following first order conditions:

$$-\frac{\partial C_i(C_i^0, \theta_g, e_{ig}, e_{-ig}|\beta)}{\partial e_{ig}} = \frac{\partial \psi_i(e_{ig}, \alpha)}{\partial e_{ig}}, \quad \forall i \in N_g^{fp}, \tag{2.5}$$

which constitutes a system of n_g^{fp} equations. For a firm belonging to a group and under a FP contract, the optimal effort level e_{ig}^{fp} is therefore conditional on the effort e_{-ig} exerted by the other members of the group:[14]

$$e_{ig}^{fp} = e_{ig}\left(C_i^0, \theta_g, \kappa_{ig} e_{-ig} \mid \beta, \alpha\right), \quad \forall i \in N_g^{fp}. \tag{2.6}$$

Solving for the n_g^{fp} equations, we obtain the unconditional effort level:

$$e_{ig}^{fp} = e_{ig}\left(C_i^0, C_{-i}^0, \theta_g, \kappa_{ig}, n_g^{fp}|\beta, \alpha\right), \quad \forall i \in N_g^{fp}. \tag{2.7}$$

Under a cost-plus contract, each operator i determines the optimal effort level that maximizes the objective function

$$\pi_i = \begin{cases} t_i^{cp} - \psi\left(e_i, \alpha\right) & \text{if } i \text{ is independent} \\ \\ t_i^{cp} - \psi\left(e_{ig}, \alpha\right) & \text{if } i \in N_g. \end{cases} \tag{2.8}$$

[13]Note that transportation networks are industries where capacity (or supply) Q is adjusted to demand levels q. As demand fluctuates during the day, the regulator determines the minimum capacity level that covers all quantities of service demanded at any moment of the day. As capacity cannot adjust instantaneously to demand levels, the minimum capacity level is always higher than demand. Hence commercial revenues are determined by q, while costs are determined by Q.

[14]Note that for each network i, the effort e_{ig}^{fp} is decreasing in the effort of the remaining networks of the group, e_{-ig}. This free-riding problem naturally arises because the cost of effort is only paid by the local network operator but benefits (at least partially) all members of the group.

In this case, since positive effort is never rewarded under a CP regime, firm i will never provide any effort, irrespective of whether it belongs to a larger industrial group. Hence we have that the optimal effort level under a cost-plus contract is given by $e_{ig}^{cp} = e_i^{cp} = 0$. Note that setting the optimal effort under a CP contract to zero is a simple normalization that we adopt for ease of exposition and tractability. We could as well assume that the operators provide the minimum effort level that guarantees to some extent the renewal of the transport concession from one period to another. There is no loss of generality because what matters in our analysis is the difference $e^{fp} - e^{cp}$.

Plugging these effort levels into the conditional cost function (2.2) yields the unconditional cost function, namely,

$$C_i^\rho(C_i^0, \theta, e^\rho|\beta) = \begin{cases} C_i^0 \times \phi\left(\theta, e_i^\rho\right) & \text{if } i \text{ is independent} \\ C_i^0 \times \phi\left(\theta_g, e_{ig}^\rho, \kappa_{ig} e_{-ig}^\rho\right) & \text{if } i \in N_g, \end{cases} \tag{2.9}$$

where $\rho = \{fp, cp\}$ refers to the type of contract. For a given firm i, equation (2.9) therefore entails two different cost structures depending on the observed regulatory regime.

2.6 Econometric specification

We now turn to the econometric specification of our cost regulation framework. In order to derive the structural cost function to be estimated, we need to assume a specific functional form for the cost function in (2.9) and the disutility of effort $\psi_i(e)$.

We assume a Cobb-Douglas specification for the cost function presented in (2.1). This specification retains the main properties desirable for a cost function while remaining tractable. Alternative more flexible specifications such as the translog function lead to cumbersome computations of the first order conditions when effort is unobservable. The primal cost function is therefore specified as:

$$C_i^0 = C_i^0(w_i, Q_i, I_i, K_i|\beta) = \beta_0 w_{L_i}^{\beta_L} w_{M_i}^{\beta_M} Q_i^{\beta_Q} I_i^{\beta_I} K_i^{\beta_K}. \tag{2.10}$$

We impose homogeneity of degree one in input prices, i.e. $\beta_L + \beta_M = 1$. In order to allow the observed cost C to deviate from the cost frontier defined by (2.10), we specify the function $\phi(\cdot)$ to be the exponential function, so that (2.2) is now specified as

$$C_i(C_i^0, \theta, e|\beta) = \begin{cases} C_i^0 \times \exp\{\theta - e_i\} & \text{if } i \text{ is independent} \\ \\ C_i^0 \times \exp\{\theta_g - e_{ig} - \kappa_{ig} \sum\limits_{j \neq i} e_{jg}\} & \text{if } i, j \in N_g. \end{cases} \tag{2.11}$$

We assume the internal cost of effort to be provided by the following convex function:

$$\psi(e) = \exp\{\alpha e\} - 1, \quad \alpha > 0, \tag{2.12}$$

with $\psi(0) = 0$, $\psi'(e) > 0$, and $\psi''(e) > 0$ and where α is a parameter to be estimated.

Using the specifications for the operating costs (2.11) and the cost of effort (2.12), we can solve the first order conditions defined in the previous section to express the optimal effort level for a network under a FP contract. We next determine the effort levels and the resulting unconditional cost functions for the different operators according to their group status and regulatory regimes.

Independent Networks

For an independent network i, the optimal effort level under a FP contract is given by the solution to (2.4) and is expressed as:

$$e_i^{fp} = \frac{1}{1+\alpha} \left(\ln(C_i^0) - \ln(\alpha) + \theta \right) \tag{2.13}$$

Recalling that $e_i^{cp} = 0$ and substituting back $e_i^{cp}$and e_i^{fp} into (2.11) allows us to obtain the final forms for the cost functions $C_i^{cp}(\cdot)$ and $C_i^{fp}(\cdot)$ to be estimated for independent networks as:

$$\ln(C_i^{fp}) = \frac{\alpha}{1+\alpha} \left[\ln(C_i^0) + \theta \right] + \frac{1}{1+\alpha} \ln(\alpha) \tag{2.14}$$

and

$$\ln(C_i^{cp}) = \ln(C_i^0) + \theta. \tag{2.15}$$

Note that equation (2.14) corresponds to the expression of the Cobb-Douglas cost function which is usually estimated, i.e. when moral hazard in the form of presence of an effort activity is not taken into account. Note also that $\lim\limits_{\alpha \to +\infty} \ln(C_i^{fp}) = \ln(C_i^{cp})$, since the effort level under a FP contract converges to 0 when the cost-reducing technology parameter α becomes infinitely large. This also translates into a lower effect of the inefficiency θ on the final costs of operator i when the cost of exerting effort is lower.

Operators Belonging to Industrial Groups

If a network i belongs to a group and is under a FP contract, it will benefit from its own cost reducing activity and from the efforts of the $n_g^{fp}-1$ remaining operators that belong to the same group and that are regulated under a fixed-price regimes as well.[15] Thus, for any industrial group g where $n_g^{fp} \geq 2$ and for any $i,\ j \in N_g^{fp}$:

$$\begin{aligned} e_{ig}^{fp} &= \frac{1}{\left(1+\alpha+\left(n_g^{fp}-1\right)\kappa_{ig}\right)} \times \left[\frac{\left(1+\alpha+\left(n_g^{fp}-2\right)\kappa_{ig}\right)}{(1+\alpha-\kappa_{ig})}\ln(C_i^0)\right. \\ &\quad \left. -\frac{\kappa_{ig}}{(1+\alpha-\kappa_{ig})}\sum_{j\neq i}\ln(C_j^0)+(\theta_g-\ln(\alpha))\right]. \end{aligned} \tag{2.16}$$

Notice how, for firm i, e_{ig}^{fp} now depends on the components defining the cost frontiers of the remaining networks of the group, $\sum\limits_{j\neq i}\ln C_j^0$. Plugging the optimal efforts and (2.16) back into the cost function (2.11) allows us to obtain the final form for the cost functions $C_{ig}^{fp}(\cdot)$ to be estimated.[16] Hence if operator i is under a FP contract and belongs to a group g where $n_g^{fp} \geq 2$, then, $\forall\, j \in N_g^{fp}$, the final form for the cost function is given by:

$$\begin{aligned} \ln(C_{ig}^{fp}) &= \frac{\alpha}{\left(1+\alpha+\left(n_g^{fp}-1\right)\kappa_{ig}\right)} \left[\frac{\left(1+\alpha+\left(n_g^{fp}-2\right)\kappa_{ig}\right)}{(1+\alpha-\kappa_{ig})}\ln(C_i^0)\right. \\ &\quad \left. -\frac{\kappa_{ig}}{(1+\alpha-\kappa_{ig})}\sum_{j\neq i}\ln(C_j^0)+\theta_g\right] + \frac{1+(n_g^{fp}-1)\kappa_{ig}}{1+\alpha+\left(n_g^{fp}-1\right)\kappa_{ig}}\ln(\alpha). \end{aligned} \tag{2.17}$$

Note how the group inefficiency θ_g is reduced by the knowledge spillovers parameter κ_{ig}. When the latter grows larger, the efforts provided in the remaining networks of the group have a larger effect on the reduction of the inefficiencies.The coefficient on the $\theta_g$parameter

[15] Recall from (2.8) that firms under a CP contract never exert any effort in equilibrium.

[16] Note that it could also be the case that firm i belongs to a group and is the only operator under a FP contract. In this case the cost function to be estimated would result in (2.14) since no other firm in the group would exert any effort. However, we do not observe such cases in our data.

is decreasing in κ_{ig} : $\frac{\partial}{\partial \kappa_{ig}}\left[\frac{\alpha}{1+\alpha+(n_g^{fp}-1)\kappa_{ig}}\right] < 0$. Likewise, the negative effect of the inefficiency parameter is reduced when the number of FP networks within the group, n_g^{fp}, increases, as operator i can benefit from the efforts of a larger number of operators. Note also that $\lim_{\kappa_{ig}\to 0} \ln(C_{ig}^{fp}) = \ln(C_i^{fp})$, as network i only benefits from its own efforts when knowledge spillovers are absent.

Recall that a network regulated under a cost-plus regime will never provide any effort, irrespective of whether it belongs to a group or not. However, if firm i belongs to group g, it will still benefit from the efforts e_{ig}^{fp} provided by the n_g^{fp} remaining operators which belong to the same industrial group and are regulated under a fixed-price regime. Thus if operator i is under a CP contract and belongs to a group g where $n_g^{fp} \geq 1$, then the effort in the n_g^{fp} remaining networks is as in (2.16) and, $\forall\ j \in N_g^{fp}$, the final form for the cost function is given by :

$$\ln(C_{ig}^{cp}) = \ln(C_i^0) + \frac{1}{\left(1+\alpha+\left(n_g^{fp}-1\right)\kappa_{ig}\right)}\left[-\kappa_{ig}\sum_{j\neq i}\ln(C_j^0) + (1+\alpha-\kappa_{ig})\,\theta_g + n_g^{fp}\kappa_{ig}\ln(\alpha)\right]. \tag{2.18}$$

Again, the effect of the group inefficiency parameter θ_g is decreasing in the knowledge spillovers parameter κ_{ig} and in the number of FP networks within the group, n_g^{fp}.

Knowledge Spillovers

We expect the different networks belonging to group g to benefit asymmetrically from the knowledge spillovers captured in the κ_{ig}, depending on several characteristics. Not every operator within a group can equally benefit from the effort exerted by other operators of the group. In particular, how much an operator will benefit from the effort of other operators will depend on how "close" they are. On the one hand, we might expect that similar networks are more likely to benefit from each other's effort. On the other hand, an operator may have more to learn and to gain from the efforts of the remaining networks in their group. Recall that the only networks that will put effort to reduce their operating costs are the networks regulated under a FP scheme. Hence, we consider that the extent to which a given operator can benefit from knowledge spillovers will depend on its similarity to the average operator under a FP contract within its group. To account for these considerations, we proxy the parameter κ_{ig} to be a function of several explanatory variables which account for the characteristics of the operator, the characteristics of the network where the service is provided and the characteristics of the group g it belongs to:

$$\kappa_{ig} = \kappa\left(\gamma_g, \delta_i, DIF^x_{i-g}\right), \tag{2.19}$$

where γ_g is a group fixed effect and δ_i is a firm fixed effect. DIF^x_{i-g} is an index which measures structural differences in the x characteristic between the observed firm i and the average firm under a FP contract $\overline{g}_{fp}$ in group g.

In our estimations, we focus on the sample containing FP networks only. That is, for a network i in period t, we estimate the cost function:

$$\ln(C^{fp}_{it}) = \xi^G_{it}\ln(C^{fp}_{igt}) + \xi^I_{it}\ln(C^{fp}_{it}) + \varepsilon_{it}, \tag{2.20}$$

where ξ^G_{it} takes value 1 if operator i belongs to one of the three main industrial groups, and 0 otherwise, while ξ^I_{it} takes value 1 if operator i is independent, and 0 otherwise. The error term ε_{it} accounts for potential measurement errors and is distributed according to a normal density function with mean 0 and variance σ^2_ε.

2.7 Data and Empirical Results

We present the estimation results of our model which are obtained by estimating the structural cost function (2.20) by maximum likelihood. We first comment the construction of the variables that enter the model.

2.7.1 Data and Variables

Different types of variables are required in order to identify our model. The cost equation calls for covariates that capture elements of the economic environment. Concerning the knowledge spillovers, we need variables that capture both group-specific and network-specific characteristics. Summary statistics are given in table 2.2, where we distinguish operators according to their group affiliation.

Estimating the Cobb-Douglas cost function requires information on the level of operating costs, the quantity of output, capital, and the input prices. Total costs C are defined as the sum of labor and material costs. Output Q is measured by the number of seatkilometers, i.e., the number of seats available in all components of rolling stock times the total number of kilometers traveled on all routes. In other words, this measure accounts for the length of the network, the frequency of the service and the size of the fleet. Note that this is also a measure of the quality of service. Capital K, which plays the role of a fixed input in our short-run cost function, is measured by the size of the rolling stock, which is the total number of seats available. Infrastructure I, which also plays the role of a fixed input, is measured by the total length of the transport network in kilometers. Since the authority owns the capital, the operators do not incur capital costs. The average wage rate w_l is obtained by dividing total labor costs by the annual number of employees. The price of materials w_m has been constructed as the average fuel price for France (published by OECD).

Estimating the knowledge spillovers requires observations on the characteristics of the operators, as well as on the features of the networks in which they operate and of the group they belong to, if any. We construct a dummy variable for each specific network and another dummy for each one of the three industrial groups (Connex, Keolis and Transdev). In order to take into account for the fact that different operators from the same group may benefit asymmetrically from knowledge spillovers, we construct a measure of the structural differences between a given firm and the average firm in the group. In particular, we define the index DIF^x_{i-g} to be a measure of the difference in the x characteristic between the observed firm i and the average firm under a FP contract in group g, $\bar{x}^{fp}_g$:

$$DIF^x_{i-g} = \frac{|x_{ig} - \bar{x}^{fp}_g|}{x_{ig}}.$$

In our estimations we consider different variables in order to calculate this index. In particular, we will focus on structural differences in the share of drivers and in the length of the network. The share of drivers is obtained by dividing the number of drivers in each network by the total labor force, which entails the bus drivers as well as engineers who are responsible for the improvement of the operator's productivity. The size of the network is measured as the total length of the transport network in kilometers. Note that this variable is also a proxy for the size of the operator. The sample that we use in our estimation is an unbalanced panel composed of 67 different networks regulated under FP contracts and contains 714 observations over the period 1987-2001.

2.7.2 Results

We turn now to the empirical results of our estimations. Table 2.3 displays the estimates of three alternative specifications. In each of them, we consider only networks regulated under FP contracts and test different explanatory variables that are used as proxies for the firm-specific knowledge spillovers within each of the industrial groups, κ_{ig}. We specify the function κ in (2.19) to have a quadratic form in all specifications. The function includes a full set of firm-specific dummy variables to control for unobserved network-specific characteristics as well as group-specific dummy variables to control for unobserved group characteristics. In other words, we specify the function in (2.19) to take the following form:

$$\kappa_{ig} = \left(\gamma_g + \delta_i + \mu DIF^x_{i-g}\right)^2. \tag{2.21}$$

For each network i belonging to group g, the dummy variable γ_g measures group g's contribution to the network's capacity to assimilate the knowledge spilled over by the other networks from the group. The dummy variable δ_i measures the network's specificities that will affect its capacity to benefit from the knowledge spillovers. These are aimed at capturing unobserved network characteristics that affect its ability to assimilate the spillovers coming from the other networks of the group. As an example, consider the firms operating the transport services in the French cities of Lille and Lyon, both belonging to the Keolis group. If the mechanic team in Lille develops a new method for repairing its buses' windshields, part of this knowledge is reached by the mechanic team in Lyon through Keolis' headquarters.[17] While the overall efficiency of both networks is similar and determined by the Keolis group, the mechanic team in Lyon may the have local characteristics which would affect its ability to fully benefit from the effort exerted by Lille's mechanic team.

Finally, we include the structural difference between a given firm and the average firm under a FP contract in the group, DIF^x_{i-g}. In specification I, it is measured using the size of the network (the total length in kilometers). We compute this index using the share of drivers in specification II.

Consider first the estimates related to the output and input variables in table 2.3. All parameter are significant at the 1% level and have the expected sign. Note that the parameter are very stable across each specification. The disutility of effort parameter, α, is also positive in both specifications, although only significant in specification II.

The effect of our similarity indexes on the knowledge spillovers parameter are positive and significant in both specifications. This result means that networks that present larger difference relative to their group (measured either in terms of drivers or length of the network) benefit to a larger extent from the efforts provided in the other networks from their group.

As already mentioned in section 5, we consider the network's intrinsic inefficiency, captured by the parameter θ_g in equation (2.17), to be group-specific. Indeed, each one of the operators belonging to group g possesses a team of engineers which is responsible for research development, quality control, maintenance, and efficiency of the network. We consider that their efficiency is determined at the group level rather than being independently determined.

[17]This example is taken from Barbosa (2009).

The estimates of the group dummy variables θ_g appear positive and highly significant. Irrespective of the specification, Keolis appears to be the most efficient group, followed by Connex and Transdev. Finally, it is interesting to note the differences in the coefficients on the group dummy variables entering the knowledge spillovers parameter, γ_g. The latter appear to be negative for each group, although it is much larger in absolute value for Connex in both specifications.

Evaluating Knowledge Spillovers

Having these estimates in hands, we are able to derive the estimated $\hat{\kappa}_{ig}$ for each network at each period. In order to test the relevance of the knowledge spillovers in our regulation model, we compute an average value of the knowledge spillovers parameter for each of the three different groups. Table 2.4 presents the results derived from each of our specifications. The estimates show statistically significant knowledge spillovers, confirming our hypothesis that operators belonging to a same group benefit from the efforts exerted by all the networks of the group. Our results also present differences across groups, with larger knowledge spillover values for the Connex and Transdev groups. Note that these two groups are the ones with the largest proportions of networks under a FP contract, highlighting the importance of this component for the knowledge spillover effects to take place.

2.8 Simulations

The resulting estimates of our structural model of regulation allow us to produce a series of counterfactual exercises. The effect of adding extra operators to a group on the final operating costs of the group members is of particular interest. Indeed, if we expect companies to benefit from knowledge spillovers, operating extra networks should help in reducing the costs of operators already in place. Another counterfactual of interest is to see what would be the effect of a merger between two groups. While such an event will undoubtedly reduce competition in the industry, it could nonetheless be beneficial if it leads to important cost reductions due to knowledge spillovers. In what follows we propose to simulate such counterfactuals in order to illustrate the potential impacts of knowledge spillovers in the French transport industry. We start by analyzing the effects of adding new operators to the existing groups.

2.8.1 Group expansion

Our model predicts that a given network i belonging to a given group g will benefit from the efforts exerted by the remaining firms in the group through the knowledge spillovers parameter κ_{ig}. We now focus our attention on the effect of increasing the number of operators in a given group, while maintaining the knowledge spillovers parameter constant. To do this, we consider a hypothetical scenario where a new operator is added to a given group and compute the cost difference resulting from that change in the group structure. We perform our simulation exercise as follows. We start by considering each operator to be the only member of its group. That is, if for example operator A originally belongs to Keolis, we consider that it is now part of a new group composed of only 1 network (namely, itself). Under this hypothetical situation, we compute the total operating cost for each operator in the sample. The next step consists in evaluating the cost change that each operator would face if a new network were added to its respective group (which, so far, consisted in only 1 network). From our cost function in (2.17), we compute the total cost associated with the operator belonging to a group composed of 2 networks. We assume the operator that is added to the group to be the representative operator of the group to which the initial network belongs. Following the example above, we would therefore add to the group of operator A (initially composed of only 1 network) an operator that is representative of the Keolis group (i.e. an operator characterized by the average values of the operators from Keolis).[18] Similarly, and using the same reasoning, we can compute the effect of adding a larger number of networks into the group. Once the simulation exercise is completed, we can easily compute the cost differential associated with the increase in group size from x to $x+p$ operators. That is, we can compute $\triangle_{xp} C_{ig}^{fp} \equiv \left(C_{ig}^{fp} \mid n_g^{fp} = x+p \right) - \left(C_{ig}^{fp} \mid n_g^{fp} = x \right)$ for $x,\, p = \{1, 2, 3, \ldots\}$.

Tables 2.5 to 2.7 present the results of this simulation exercise using the estimates derived from specification II. [19] Results show significant costs reductions from being linked to a larger number of networks. In particular, the cost savings are increasing importantly with the number of operators that are added to the group. Note also that the effect of an additional network varies in function of the initial group size. Although to a small extent, the cost reduction associated with an extra operator is decreasing in the size of the group. Finally, note that the different groups benefit from the inclusion of additional operators to different extents. In particular, and in accordance with the knowledge spillovers values presented in table 2.4, Connex and Transdev benefit to a larger extent from the inclusion of extra networks into their group.

[18] Note that for any operator i that belongs to group g, the variable DIF_{i-g}^{x} is unaffected by the addition of an operator that is representative of the existing firms already in the group. It follows that the knowledge spillovers parameter κ_{ig} is unaffected by such a change.

[19] The simulations based on the estimates derived from specifications I show similar results.

2.8.2 Merger

As already mentioned, industrial groups of urban transport have a long history of mergers in France. The last merger that was witnessed in the French transport industry occurred on March 3rd 2011 and involved *Veolia Transport* (the former Connex) and Transdev, which gave birth to *Veolia Transdev.* Our model allows us to simulate such a merger and to evaluate the potential gains in costs for the merging groups, namely Connex and Transdev.[20] Several assumption must be made on the post merger outcomes regarding our parameters. We first assume that the inefficiency level of the group resulting from the merger (Veolia Transdev in our example) will take the value of the most efficient merging group. In other words, we assume that the less efficient group is absorbed by the most efficient one. In all of our estimations above, the estimated parameters $\hat{\theta}_g$ show that Connex is the most efficient group of the two since $\hat{\theta}_{Transdev} > \hat{\theta}_{Connex}$. We therefore assume that $\theta_{VeoliaTransdev} = \theta_{Connex}$ in our simulations. Similarly, we assume that, after the merger, the group-specific capacity to transmit knowledge will be the highest of the two merging groups. In all of our estimations results, the estimated parameters $\hat{\gamma}_g$ show that Transdev has the highest capacity to transmit knowledge among its operators since $\hat{\gamma}_{Transdev} > \hat{\gamma}_{Connex}$. We therefore assume that $\gamma_{VeoliaTransdev} = \gamma_{Transdev}$ in our simulations. Finally, note the difference between this simulation exercise and the one we carried in the previous section. In the latter, we computed the cost reduction associated with an increase in the group size while maintaining the knowledge spillover parameter κ constant. Here, κ will also change as a result of the merger since the operators entering the new group are not necessarily representative of the ones already in place within the group.[21] Table 2.8 presents the results of the merger simulation exercise using the estimates derived from each of our specifications. For each one of the merging groups, each cell in the table presents the average percentage change in the operators' costs following the merger. The results are very similar across each specification and show important gains in costs from the merger. In particular, cost reductions are very important for operators initially belonging to Connex, sometimes twice as large as the cost reductions for operators initially belonging to Transdev. This last fact is perhaps not surprising for two reasons. First, according to our results in table 4, Connex is the group that benefits the most, on average, from knowledge spillovers. Second, Transdev is the group with the largest number of networks regulated under FP contracts (see table 2.1). It follows that, after the merger, the operators initially belonging to Connex see their number of FP networks increase much more relative to operators initially belonging to Transdev. For instance, in our first year of data (1987), Connex was present in 5 different networks regulated

[20]Note, however, that our simulation does not correspond with the same years in which the actual merger occurred. Since we simulate the merger with our entire sample that covers the period 1987-2001, it is as if the merger had occurred in 1987.

[21]Recall that the simulation exercise in section 2.8.1 was realized by adding representative operators to a group, maintaining constant the DIF^{x}_{i-g}variable throughout the exercise (see footnote 18).

under a FP contract while Transdev was present in 15 such networks. After a merger, the operators initially belonging to Transdev would therefore see the number of FP networks in their group increase by a third (from 15 to 20) while operators initially belonging to Connex would see the number of such networks quadruple (from 5 to 20). It is however important to highlight that our model does not capture the fact that not all of the 15 FP networks from Transdev will equally contribute to Connex's cost savings when the merger takes place (nor will all of the 5 FP networks from Connex equally contribute to Transdev's cost savings). In particular, one might expect merging parties to show decreasing returns to the number of merging networks with a given set of characteristics. To illustrate this point, consider table 2.9 which presents a measure of the congestion levels of the different FP networks involved in our merger simulation.[22] While Connex will benefit from the efforts of all 15 networks from Transdev after the merger, our previous results have suggested that it will benefit from them based on their characteristics rather than based on their number. For instance, table 2.9 indicates that Connex operates a single network with a congestion level of around 850 inhabitants per km of line. Transdev, on the other hand, is operating 5 different networks with congestion levels within this range (between 600 and 950 inhabitants per km of line). It therefore seems natural to think that the relevant knowledge that Connex would benefit from would come for these 5 networks as a whole. Given that our model and simulation does not explicitly take this consideration into account, the results from our simulations should be taken as an upper bound on the potential gains that could follow from the merger and should be interpreted with caution. In the same line, one must still take into account that the main trade-off involved in a merger evaluation is between potential efficiency gains resulting from consolidation versus the potential increases in price and potential deterioration of the quality of the transportation service. It is therefore important to highlight that our simulations only allow us to evaluate one part of this trade-off, namely the potential gains in costs that would follow from the merger. From that perspective, our results confirm the importance of considering the structure of the industrial transportation groups at the time of evaluating mergers in the French transportation industry.

2.9 Conclusion

In this paper we identify and measure the relevance of knowledge spillovers in urban transport regulation. We take advantage of a specific feature of the French urban transport industry, namely that about eighty percent of the operators that provide the transport services in each city are owned by three large industrial groups. The transport services provided in different networks by operators belonging to the same industrial group are therefore essentially

[22]These are networks that were operated by either Connex or Transdev in 1987, the first year of our data, and our congestion measure is computed as the ration of the population over the size of the network in kilometers.

provided by the same firm. On top of that, the activity of every network is regulated by a local authority within a specific regulatory framework. The latter takes the form of a written contract that can be, in practice, either cost-plus or fixed-price. While effort to compensate technological inefficiencies is not rewarded under a cost-plus contract, fixed-price contracts provides powerful incentives to reduce operating costs.

When operators belong to a same group, the new methods and procedures that they develop can potentially be used by the entire company. We build and estimate a structural cost regulation model with asymmetric information that includes knowledge spillovers resulting from operating different networks. By focusing our analysis on operators regulated under fixed-price contracts, we ask whether their linkage through a larger group helps them further reduce their operating costs.

Our results show statistically significant knowledge spillovers, confirming the existence of relevant knowledge spillovers in the French public urban transportation industry. Furthermore, several simulation exercises derived from our estimates show that operators gain significantly from being linked to a larger number of networks within a group. In particular, the cost reductions following the addition of new operators into a group is increasing in the number of networks added. Finally, the simulation of a merger between the Connex and Transdev groups, as actually occurred on March 3rd 2011, show important costs reductions for the operators involved, between 24% and 49%. Our results therefore provide evidence on the importance of taking knowledge spillovers into account when evaluating the economic effects of mergers.

Table 2.1: Characteristics of regulatory contracts and group affiliation.

Variable Name	Frequency	Percent
Networks	87	
Observations corresponding to FP contract		61.0
Belongs to a group if under FP contract		83.1
Changes in contract type	23	
Changes from CP to FP	18	
Changes of operator	5	
Changes from group operator to indep. operator	2	
Changes from indep. operator to group operator	1	
changes from group operator to group operator	2	
Operator belongs to a group		78.6
FP if operator belongs to a group		64.5
Operator is independent		21.4
FP if operator is independent		48.2
Operator belongs to Keolis	410	
FP if operator belongs to Keolis		51.7
Operator belongs to Transdev	269	
FP if operator belongs to Transdev		90.7
Operator belongs to Connex	241	
FP if operator belongs to Connex		56.9

Note: CP refers to cost-plus contracts and FP refers to fixed price contracts

Table 2.2: Summary statistics by type of operator

		Type of operator			
		Belongs to group		Independent	
Name	Variable	Mean	Std dev.	Mean	Std dev.
Cost (Euros)	C	18157.23	26883.30	6473.55	5261.45
Revenue (Euros)	$R(q)$	8400.05	13322.25	2765.65	2299.52
Production (Seat-kilometers)	Q	579178.70	748774.50	240941.80	181420.00
Wage (Euros)	w_L	29.66	5.72	29.11	6.36
Price of materials (Index)	w_M	1.17	0.20	1.18	0.20
Size of the network (Kil.)	$length$	256.05	223.84	153.51	87.49
% of drivers in the labor force	$Drive$	0.72	0.08	0.74	0.07

Note: Group refer to operators belonging to either Keolis, Transdev or Connex.

Table 2.3: Structural Estimation Results

Name	Parameter	I	II
Constant		-3.648^{***}	-3.684^{***}
		(0.112)	(0.109)
Connex	θ_{Connex}	0.545^{***}	0.522^{***}
		(0.063)	(0.052)
Keolis	θ_{Keolis}	0.418^{***}	0.374^{***}
		(0.044)	(0.046)
Transdev	$\theta_{Transdev}$	0.630^{***}	0.607^{***}
		(0.057)	(0.053)
Wage	β_L	0.279^{***}	0.273^{***}
		(0.041)	(0.034)
Production	β_Q	1.042^{***}	1.059^{***}
		(0.199)	(0.148)
Infrastructure	β_I	0.124^{***}	0.145^{**}
		(0.023)	(0.021)
Cost of effort	$\ln(\alpha)$	1.719	1.624^{**}
		(1.053)	(0.708)
Connex	γ_{Connex}	-0.242^{**}	-0.225^{***}
		(0.112)	(0.070)
Keolis	γ_{Keolis}	-0.119^{*}	-0.068^{**}
		(0.069)	(0.031)
Trans	γ_{Trans}	-0.121	-0.076^{*}
		(0.074)	(0.039)
Dif Length	DIF_{i-g}^{Len}	0.008^{**}	
		(0.003)	
Dif Drivers	DIF_{i-g}^{Dri}		0.072^{***}
			(0.024)
Stand. Dev. error	σ_ϵ	0.102^{***}	0.104^{***}
		(0.003)	(0.003)
Firms fixed effects	δ_i	*yes*	*yes*
Number of observations		714	714

Note: The sample contains networks regulated under a FP contract. Standard errors in parenthesis. ***: Significant at 1%, **: Significant at 5%, *: Significant at 10%.

Table 2.4: Average Knowledge Spillovers by Group

	Specification	
Group	I	II
Connex	0.017	0.015
	(0.001)	(0.001)
Keolis	0.007	0.005
	(0.000)	(0.000)
Transdev	0.014	0.012
	(0.000)	(0.000)

Note: Each cell represents the average knowledge spillovers of the corresponding group, computed as $\bar{\kappa}_g = \frac{1}{n_g^{fp}} \sum_{i \in g} \hat{\kappa}_{ig}$. Standard errors are in parenthesis.

TABLE 2.5: Percentage change in total costs from adding new operators for Connex

Going from x operators...	... to $x+p$ operators in the Connex group: 2	3	4	5	6	7	8	9	10
1	−4.51	−8.58	−12.28	−15.65	−18.74	−21.58	−24.19	−26.62	−28.86
	(0.317)	(0.587)	(0.821)	(1.023)	(1.120)	(1.353)	(1.488)	(1.608)	(1.715)
2	-	−4.40	−8.39	−12.03	−15.35	−18.40	−21.21	−23.80	−26.21
	-	(0.306)	(0.568)	(0.796)	(0.994)	(1.167)	(1.319)	(1.454)	(1.573)
3	-	-	−4.30	−8.22	−11.79	−15.06	−18.07	−20.85	−23.42
	-	-	(0.295)	(0.550)	(0.772)	(0.966)	(1.136)	(1.287)	(1.420)
4	-	-	-	−4.21	−8.04	−11.56	−14.78	−17.75	−20.50
	-	-	-	(0.285)	(0.533)	(0.749)	(0.939)	(1.107)	(1.255)
5	-	-	-	-	−4.12	−7.88	−11.33	−14.51	−17.44
	-	-	-	-	(0.276)	(0.516)	(0.728)	(0.914)	(1.078)
6	-	-	-	-	-	−4.03	−7.72	−11.11	−14.25
	-	-	-	-	-	(0.267)	(0.501)	(0.707)	(0.889)
7	-	-	-	-	-	-	−3.94	−7.57	−10.90
	-	-	-	-	-	-	(0.258)	(0.486)	(0.687)
8	-	-	-	-	-	-	-	−3.863	−7.419
	-	-	-	-	-	-	-	(0.250)	(0.472)
9	-	-	-	-	-	-	-	-	−3.784
	-	-	-	-	-	-	-	-	(0.243)

Note: Each cell in the table presents the average percentage change in costs following the addition of p networks to a group in which x operators are already present: $\triangle_{xp} C_{ig}^{fp} \equiv \left[\frac{\left(C_{ig}^{fp} \mid n_g^{fp}=x+p\right) - \left(C_{ig}^{fp} \mid n_g^{fp}=x\right)}{\left(C_{ig}^{fp} \mid n_g^{fp}=x\right)} \right] \times 100$, using equation (2.17) in the text. E.g. the cell corresponding to row 2 and column 7 give the percentage change in the cost of an operator that goes from belonging to a group of 2 operators to a group of 7 operators. Standard errors in parenthesis. Estimates are computed from the results obtained in specification II.

TABLE 2.6: Percentage change in total costs from adding new operators for Keolis

		... to $x+p$ operators in the Keolis group								
		2	3	4	5	6	7	8	9	10
Going from x operators...	1	−1.62	−3.17	−4.66	−6.09	−7.47	−8.79	−10.06	−11.29	−12.48
		(0.112)	(0.216)	(0.313)	(0.403)	(0.487)	(0.565)	(0.639)	(0.708)	(0.773)
	2	-	−1.60	−3.14	−4.62	−6.04	−7.40	−8.72	−9.98	−11.21
		-	(0.110)	(0.213)	(0.308)	(0.397)	(0.480)	(0.557)	(0.630)	(0.699)
	3	-	-	−1.59	−3.11	−4.58	−5.99	−7.34	−8.65	−9.90
		-	-	(0.109)	(0.209)	(0.303)	(0.391)	(0.473)	(0.550)	(0.622)
	4	-	-	-	−1.57	−3.08	−4.54	−5.93	−7.28	−8.58
		-	-	-	(0.107)	(0.206)	(0.298)	(0.385)	(0.466)	(0.542)
	5	-	-	-	-	−1.56	−3.06	−4.50	−5.88	−7.22
		-	-	-	-	(0.105)	(0.203)	(0.294)	(0.379)	(0.459)
	6	-	-	-	-	-	−1.54	−3.03	−4.46	−5.83
		-	-	-	-	-	(0.103)	(0.199)	(0.289)	(0.374)
	7	-	-	-	-	-	-	−1.53	−3.00	−4.42
		-	-	-	-	-	-	(0.101)	(0.196)	(0.285)
	8	-	-	-	-	-	-	-	−1.52	−2.98
		-	-	-	-	-	-	-	(0.100)	(0.193)
	9	-	-	-	-	-	-	-	-	−1.50
		-	-	-	-	-	-	-	-	(0.098)

Note: Each cell in the table presents the average percentage change in costs following the addition of p networks to a group in which x operators are already present: $\triangle_{xp} C_{ig}^{fp} \equiv \left[\frac{\left(C_{ig}^{fp} | n_g^{fp}=x+p\right) - \left(C_{ig}^{fp} | n_g^{fp}=x\right)}{\left(C_{ig}^{fp} | n_g^{fp}=x\right)}\right] \times 100$, using equation (2.17) in the text. E.g. the cell corresponding to row 2 and column 7 give the percentage change in the cost of an operator that goes from belonging to a group of 2 operators to a group of 7 operators. Standard errors in parenthesis. Estimates are computed from the results obtained in specification II.

TABLE 2.7: Percentage change in total costs from adding new operators for Transdev

		... to $x+p$ operators in the Transdev group								
		2	3	4	5	6	7	8	9	10
Going from x operators...	1	−3.77	−7.29	−10.58	−13.66	−16.54	−19.25	−21.80	−24.20	−26.46
		(0.151)	(0.288)	(0.412)	(0.525)	(0.627)	(0.720)	(0.804)	(0.881)	(0.952)
	2	-	−3.71	−7.18	−10.42	−13.47	−16.32	−19.00	−21.53	−23.90
		-	(0.148)	(0.282)	(0.404)	(0.514)	(0.615)	(0.706)	(0.790)	(0.866)
	3	-	-	−3.66	−7.08	−10.28	−13.28	−16.10	−18.76	−21.26
		-	-	(0.145)	(0.276)	(0.395)	(0.504)	(0.603)	(0.693)	(0.776)
	4	-	-	-	−3.60	−6.97	−10.13	−13.10	−15.89	−18.52
		-	-	-	(0.142)	(0.270)	(0.387)	(0.494)	(0.592)	(0.681)
	5	-	-	-	-	−3.546	−6.87	−9.99	−12.92	−15.68
		-	-	-	-	(0.139)	(0.265)	(0.380)	(0.485)	(0.581)
	6	-	-	-	-	-	−3.49	−6.77	−9.85	−12.75
		-	-	-	-	-	(0.136)	(0.259)	(0.372)	(0.475)
	7	-	-	-	-	-	-	−3.44	−6.67	−9.71
		-	-	-	-	-	-	(0.133)	(0.254)	(0.365)
	8	-	-	-	-	-	-	-	−3.39	−6.58
		-	-	-	-	-	-	-	(0.130)	(0.249)
	9	-	-	-	-	-	-	-	-	−3.34
		-	-	-	-	-	-	-	-	(0.127)

Note: Each cell in the table presents the average percentage change in costs following the addition of p networks to a group in which x operators are already present: $\triangle_{xp} C_{ig}^{fp} \equiv \left[\frac{\left(C_{ig}^{fp}|n_g^{fp}=x+p\right)-\left(C_{ig}^{fp}|n_g^{fp}=x\right)}{\left(C_{ig}^{fp}|n_g^{fp}=x\right)}\right] \times 100$, using equation (2.17) in the text. E.g. the cell corresponding to row 2 and column 7 give the percentage change in the cost of an operator that goes from belonging to a group of 2 operators to a group of 7 operators. Standard errors in parenthesis. Estimates are computed from the results obtained in specification II.

Table 2.8: Percentage changes in total costs after merger

	Specification	
Group	I	II
Connex	-49.28	-43.18
	(4.632)	(4.398)
Transdev	-23.88	-24.02
	(0.779)	(0.763)

Note: Each cell gives the average percentage change in the total costs of the operators of the corresponding merging group. Standard errors are in parenthesis.

Table 2.9: Congestion rate for the networks operated by Connex and Transdev in 1987

	Group	
	Transdev	Connex
Le Creusot	354.7	
Thionville	393.0	
Longwy	602.3	
Valenciennes	668.7	
Orleans	790.2	
Montpellier	860.5	
Bayonne	944.6	
Toulouse	1068.5	
Ajaccio	1212.5	
Limoges	1340.6	
Metz	1341.1	
Agen	1465.0	
Grenoble	1480.2	
Macon	1904.4	
Strasbourg	2136.4	
Niort		847.2
Dunkerque		1336.3
Nice		1350.3
Cannes		1562.5
Aubagne		2116.1

Note: Each cell gives the network congestion rate as the total number of inhabitants per km of line in 1987.

Chapter 3

Digital Music Consumption on the Internet: Evidence from Clickstream Data

3.1 Introduction

In the last decade, digitization has dramatically affected most of the media industries. Digital technologies have allowed to drastically reduce the costs of copying and disseminating information. In the case of the music industry, these costs reductions have lead to major gains for consumer who can now easily enjoy and benefit from a wider range of products at a minimal cost. Music producers, on the other hand, have for many years feared the advent of digitization, and in particular piracy, in which they saw a major threat to their revenues. Understanding how technological change and digitization have affected the music industry as a whole is important in order to assess its effects on welfare.

In order to promote innovation and maximize welfare, copyright protection trades off the costs of limiting access to a creative work (e.g. a song) against the benefits of providing incentives to create it (Landes and Posner, 1989). By effectively weakening copyright protection, music piracy may wreak havoc with the objective of maximizing society's welfare. Understanding the effects of piracy is therefore of major importance from a public policy perspective. The increase of illegal music consumption is worrisome because it could lead to a decrease in music producers' revenues and consequently to a reduction of the supply of innovative music. A necessary condition for this claim to hold true, however, is that legal and illegal consumption of music must be reasonably close substitutes. In this case, songs obtained via unauthorized channels effectively depress sales since they would otherwise have been purchased. Illegal music consumption is therefore only a potential threat to artistic production inasmuch as it displaces legal consumption.[1]

The impact of music piracy on legal sales of music has been studied extensively in the empirical literature, focusing mainly on legal music sales in the form of physical CDs. Most studies find that piracy harms revenues, with estimated sales displacement rate far below one. That is, music consumers are found to substitute legal music consumption for illegal music consumption, but much of what is consumed illegally would not have been purchased if piracy was not available. There is a rather clear consensus on the negative effects of online piracy on the off-line physical sales of recorded music. This naturally leads to a concern about the potential negative effect that piracy could have on the flow of music to be delivered to market.[2]

Since the launch of the iTunes music store in 2003, the availability to purchase legal digital songs changed individuals' music consumption alternatives. Instead of having to buy a whole CD, the alternative to downloading any particular digital song illegally is now to purchase it in MP3 format. As emphasized in Waldfogel (2010), the appearance of file-sharing and downloading technology might have different effects on sales, depending on whether the legal option is a 12-song CD or à la carte songs. Consider an individual interested in a few songs of a given artist. While she may not consider buying the entire album (which also contains *unknown* songs) when offered the possibility to freely download these specific songs, she might nevertheless be willing to pay for them individually. The effect of downloading on individual songs and albums may therefore be different, and one can easily imagine a circumstance in which file-sharing would hurt album sales more than it hurts song sales.

The empirical literature on music piracy has paid much less attention to the effect of illegal music consumption on the legal sales of digital music. In this paper, we ask whether online music consumers perceive illegal digital music as a close substitute to legal digital music consumption. We therefore revisit the question of sales displacement in the digital era, adding evidence to a fundamental debate in the economics of copyright. Second, we analyze how online music streaming affects the purchases of digital music, a question that has received very little attention in the empirical literature thus far. Finally, a key contribution to this paper is the originality of its dataset, which helps us circumvent the inherent difficulties in studying illegal behavior such as file-sharing. Our approach relies on a novel dataset that enables us to follow a large sample of Internet users and their online behavior in five EU

[1]Note that this is only a necessary, but not sufficient condition. One must in particular take the effects of digitization on the costs of production into account. These could indeed potentially benefit both producers and consumers alike and offset the potentially negative effect of piracy (Waldfogel, 2012c). See Oberholzer-Gee and Strumpf (2010) for an extended discussion on file-sharing, copyright protection and the incentives to create, market and distribute new works.

[2]The empirical litterature has nevertheless failed to identified any negative effect of digitization on the supply of music brought to market (Handke, 2012; Waldfogel, 2011, 2012a,b).

countries during 2011. For each of the individuals in our sample, we observe both information on demographic characteristics and on the webpages visited during the year. This allows us to identify specific visits on websites related to music consumption, both legal and illegal. Tracking individual online behavior also allows us to construct other variables reflective of otherwise unobserved characteristics, such as taste for music. All of these features, combined with the panel structure of our data, allows us to control for many forms of unobserved heterogeneity that would otherwise jeopardize the identification of a causal effect of illegal downloading (and legal online streaming) on the legal purchases of digital music.

Perhaps surprisingly, our results present no evidence of digital music sales displacement. While we find important cross country differences in the effects of downloading on music purchases, our findings suggest a rather small complementarity between these two music consumption channels. It seems that the majority of the music that is consumed illegally by the individuals in our sample would not have been purchased if illegal downloading websites were not available to them. The complementarity effect of online streaming is found to be somewhat larger, suggesting a stimulating effect of this activity on the sales of digital music.

Taken at face value, our findings indicate that digital music piracy does not displace legal music purchases in digital format. This means that although there is trespassing of private property rights (copyrights), there is unlikely to be much harm done on digital music revenues. This result, however, must be interpreted in the context of a still evolving music industry. It is in particular important to note that music consumption in physical format has until recently accounted for the lion's share of total music revenues.[3] If piracy leads to substantial sales displacement of music in physical format, then its effect on the overall music industry revenues may well still be negative.

We cannot draw policy implications at the industry-wide level, as our analysis is only confined to the digital segment of the music industry. Nonetheless, digital music revenues to record companies are growing substantially. They increased more than 1000% during the period 2004-2010, and grew 8% globally in 2011 to an estimated US$5.2 billion, reflecting the importance of digitization in the music industry (IFPI, 2011, 2012).[4] From that perspective, our findings suggest that digital music piracy should not be viewed as a growing concern for copyright holders in the digital era. In addition, our results indicate that new music consumption channels such as online streaming positively affect copyrights owners.

The remainder of the paper is organized as follows. Section 3.2 summarizes the underlying theory as well as the relevant literature on the subject. It presents the results of the main

[3]In the case of the UK, it is indeed only in the first quarter of 2012 that sales from digital sales surpassed sales of traditional CDs and records for the first time, see `http://www.guardian.co.uk/media/2012/may/31/digital-music-spending-bpi`.

[4]This compares to growth of 5% in 2010 and represents the first time the year-on-year growth rate has increased since IFPI started measuring digital revenues in 2004 (IFPI, 2012).

empirical studies on the effects of piracy on record sales. Section 3.3 presents the data and the different variables used in the estimation. Section 3.4 presents our empirical approach and the results of our estimations. Finally, section 3.5 discusses the results and concludes.

3.2 Theory and Related Literature

Economic theory does not provide a clear prediction for how illegal downloading should affect legal music consumption.[5] The crucial point is to know whether illegal consumption (the downloading of an album or a song) would have been converted into legal consumption (the purchase of that same album or song) in the absence of illegal consumption channels. If the albums consumed through illegal channels are valued above their price by the consumer, then there is indeed sales displacement: the consumer would have bought the album had she not downloaded it. If, however, the consumer's valuation is below the album's price, then no sales displacement occurs: the consumer would not have bought the album had she not downloaded it. Given the heterogeneity of consumers, the willingness to pay will be above the market price for some and below the market price for others, leading to an average displacement rate between zero and one. Considering this simple static configuration, it follows that the availability of illegal music consumption channels unambiguously increases welfare.[6] All instances of sales displacement will simply convert some of the producers' revenues into consumers surplus, while illegal consumption from low valuation individuals (individuals with valuations lower than the price) will increase consumer surplus without hurting revenues (Rob and Waldfogel, 2006; Waldfogel, 2010).

Illegal music consumption could also, in theory, stimulate legal music consumption. Since music is an experience good, file sharing can allow consumers to sample specific songs or albums which can inform them on what to buy. Similarly, the sampling of a specific song may stimulate individual demand for other songs by the same artist (Shapiro and Varian, 1999; Peitz and Waelbroeck, 2006; Belleflamme and Peitz, 2010).

Given all these considerations, the question of whether consumers' ability to illegally obtain free recorded music displaces legal music consumption remains an empirical one. An important and still growing amount of research has explored this question, using different data sources and different approaches. The reasons for the inherent difficulty in measuring the effect of illegal downloading on legal music sales are twofold. First, downloading is an illegal behavior, which renders is measurement difficult. It is therefore not easy to obtain data on unpaid consumption nor to link it to data on paid music consumption. Second, assuming

[5] We will use the terms downloading and file sharing interchangeably to refer to illegal music consumption in the remainder of the text.

[6] Note that this leaves out the dynamic considerations of the issue.

that such data is available, identifying the causal effect of downloading on legal purchases is made difficult by the non-experimental nature of the data. The main challenge to overcome is the existence of unobserved heterogeneity that renders the downloading variable potentially endogenous.

Empirical researchers have pursued different types of strategies to come around these difficulties. A first set of papers uses time series data at the geographic level in order to compare the music sales levels in different location over time. The main idea is then to ask whether places with higher levels of piracy (typically proxied by measures of Internet broadband penetration) present lower levels of sales. Some studies following this approach include Hui and Png (2003), Peitz and Waelbroeck (2004), Zentner (2009) and Liebowitz (2008), all of which find some displacement of physical music purchases by illegal downloads.

A second category of papers uses product level data (i.e. record data) to see whether records that are downloaded more are purchased more or less. Some researchers have used natural experiments to identify the causal effect of piracy on sales. Danaher et al. (2012) use the HADOPI graduated response law in France as an exogenous shock and compare iTunes music sales in France to sales in a set of other European countries. They find that HADOPI caused a 22.5% increase in song sales and a 25% in album sales relative to sales in the control group, which is consistent with Internet piracy displacing legal iTunes sales.

Often lacking such natural experiments, others researchers have used an instrumental variable approach to deal with the endogeneity of piracy. In a widely cited paper, Oberholzer-Gee and Strumpf (2007) construct a weekly panel of album sales and illegal downloads. They use the number of German secondary school students who are on holidays in specific weeks as instruments for downloads and find that file sharing has an effect on sales that is statistically indistinguishable from zero. Ina recent study, Hammond (2012) focuses on pre-release file sharing, in which file sharers download sound recordings that are not yet publicly available. Using instrumental variables to deal with the endogeneity of file sharing, he finds that the causal effect of file sharing of an album on its sales is essentially zero.

The third approach used in the empirical literature is to use individual-level (survey) data, asking whether consumers who engage in illegal music consumption engage in more or less paid consumption.[7] When using cross-sectional data, the presence of unobserved heterogeneity across individuals (in particular music taste) is an important obstacle to the identification of the causal effect of downloading on legal purchases. Using a survey administered to U.S. university students in 2003, Rob and Waldfogel (2006) rely on an instrumental variable approach with access to broadband as a source of exogenous variation in downloading. They find that each album download reduces purchases by about .2 in their sample. Zentner (2006) follows a similar approach using a cross-section of 15000 European individuals in

[7]See also Rob and Waldfogel (2007) and Bai and Waldfogel (2012) for the case of movie piracy.

2001. Instrumenting for piracy using Internet connection speed as well as levels of Internet sophistication, he finds that people who self-report downloading music are also less likely to have recently purchased music.

As highlighted by Smith and Telang (2012), there are two main interpretation challenges that arise when using a survey-based approach. First, the conclusions are, inevitably, tied to the chosen sample. This is problematic if one believes that the sample is not representative of the overall population of interest. Although a study based on a sample of university students may still lead to insightful results, one cannot generalize them to a population other than the one of college students. Second, surveys can be affected by inaccurate recall or obfuscation from the respondents. In particular, individuals my voluntarily under- or over- represent their actual purchase or illicit behavior.

Although some specific papers fail to find evidence of sales displacement, the emerging consensus on the effect of piracy is that unpaid consumption depresses music sales. The displacement effects found are, however, typically less than 1, indicating that much of what is downloaded would not have been purchased in the absence of illegal consumption channels.

With the exception of Danaher et al. (2012), all of the above studies use data drawn from times in which the standard legal option offered by the music industry was a physical CD. Using two surveys of undergraduate college students, Waldfogel (2010) analyzes the effect of piracy when legal digital options are available. He finds, however, that the rate of sales displacement in both samples is similar to the one observed before legal digital options were available. More specifically, each additional downloaded song is found to reduce paid consumption by between a third and a sixth of a legally purchased song. A recent study shows results that go in the opposite direction. Using survey data on a sample of 2000 French individuals, Bastard et al. (2012) find that while piracy has a negative effect on the probability to purchase music in CD format, it has a positive effect on the probability of downloading music legally. Hence legal music downloading and piracy are complements rather than substitutes in their sample. Finally, DangNguyen et al. (2012) is, to our knowledge, the only empirical study that analyzes the effect of streaming on music purchases. Based on survey data on 2000 French consumers, they find that consuming music as streams has no significant effect on CDs purchases but is a complement to buying music online. Our findings are in line with the results of this recent research.

The limited number of studies analyzing the effect of piracy on sales in times when consumers are offered legal digital alternatives therefore offers rather mixed results. Given the importance of copyright protection in the promotion of innovation and welfare, this scarce amount of evidence calls for further research on that crucial question.[8]

Several features of our data allow us to contribute to the existing literature presented above. First, we have access to a sample of Internet users that is representative of the online population in five different European countries in terms of gender and age. As opposed to studies based on specific samples (e.g. college students), the results of our analysis therefore need not be restricted to a particular part of the population. Second, contrary to studies based on individual surveys, our data does not rely on subjective assessment from Internet users but on actual consumption patterns. In particular, browsing activity of Internet users allow us to construct many specific variables (such as proxies for interest in music) that will allow us to control for otherwise unobserved individual characteristics. Finally, the panel dimension of our data will allow us to further control for time invariant unobservables. We now turn to a detail presentation of our data.

3.3 Data and Variables

3.3.1 The Data

The original data on which we rely comes from Nielsen NetView, which is Nielsen's Internet audience measurement service. It uses metered measurement of representative panels of Internet users to track usage across websites and digital applications. The service also reports demographic information on the Internet users. Nielsen initially aims at gathering a sample that is representative of the overall Internet audience at home for people aged at least 2 years old and with access to Internet in each country.

The Nielsen Clickstream Data provides a very rich set of information on both consumers' demographic characteristics and online behavior. The sample that we have available contains information on 5000 individuals for each of the five largest European economies: France, Germany, Italy, Spain and the UK. First, we have access to information about the socioeconomic characteristics of each user. In particular, we observe gender, age, education, occupation, household income, household size, presence of children in the household and region of residence. Second, the original database contains all the clicks of each of the 25,000 Internet user for the period going from January 1st, 2011 to December 31st, 2011. For each of these clicks, we observe the URL of the webpage visited and the time at which it was visited, the duration of time that the webpage is viewed and a classification of the webpage according to its content. There is a total of 15 different categories, which contain a total of 83 subcategories.

[8] Some empirical studies have looked at the effect of piracy on the legal consumption of other digital content. See for example Danaher et al. (2010) for the case of television content and Danaher and Smith (2013) for the case of movie digital sales.

The main task that needed to be carried out was the identification and classification of websites related to music consumption. By that we mean websites whose direct purpose is the listening of music. These can take several forms, which constitute our different categories of music consumption: music downloading, music streaming, music-video streaming, and radio. The downloading and streaming categories can further be divided into legal and illegal websites.[9] We will restrict our analysis to the sample of individuals who consume music through legal purchases, illegal downloading or legal streaming, meaning that we leave out the individuals that never visited one of these specific music consumption websites during 2011. We consider individuals aged between 10 and 75. The focus of our study is to analyze the relationship between several channels of digital music consumption, and in particular on the causal effect of illegal downloading and legal streaming on legal purchases of digital music. We therefore focus on individuals that are involved in either one of these three activities. After dropping individuals with missing values, we are left with a total of 16,290 individuals.[10]

3.3.1.1 Websites

We identified a total of 2,759 music consumption related websites in our database, which amounted to a total of 5,054,389 clicks during 2011. The classification of websites was done by going on the mostly visited ones and checking their purpose and origin. We decided to restrict our attention to the websites that had received more than 300 clicks during 2011, leaving us with a total number of 779 websites to check manually.[11] Since the distribution of clicks is very concentrated on specific websites, our selection of websites covers 4,956,243 clicks, i.e. 98% of the total clicks.

It is important to note that we are only able to observe the number of clicks on a given website and that we do not have a precise description of the individual behavior for each click. Rather than measuring actual consumption or purchases, our data therefore gives a measure of the propensity to consume music. We believe, however, that this is still a good approximation to actual consumption. We see no specific reason for which an individual would go on a music-consumption website with other purposes than to consume music. While this is especially true for illegal downloading and legal streaming websites, the proportion of clicks that lead to a purchase for visits on legal purchasing websites could be expected to be

[9] The observations on illegal music streaming websites are quite scarce in our data set.

[10] Missing values come mainly from the demographic variables, where some individuals failed to respond. Note that our panel can be constructed at pretty much any time dimension. We have constructed one version at the week level (52 observations per individual and one version at the month level (12 observations per individual). As expected the weekly version contains many more 0 values than the monthly version.

[11] Notice that the total number of visits (clicks) accounts for the overall database. Less than 300 clicks therefore means less than 1 visit a day among 25,000 individuals in 5 different countries, a rather small number.

lower due to simple browsing activity. Still, we believe that this (possibly) lower fraction of purchasing-clicks does not reflect any particular individual characteristic. In particular, we do not expect individuals to go window-shopping on legal purchasing websites in order to illegally download after their visit. First, information on specific albums, songs or artists can be found on other music-specific websites, so it is not clear why consumers should use legal purchasing websites for such purposes. Second, we believe information on songs' prices to be almost perfectly known to consumers before they go on legal purchasing websites, ruling out visits solely related to price information seeking. These features of our data suggest that our variables of interest are measured with some errors. Our coefficient may therefore suffer from attenuation bias, which would potentially bias them toward zero.

Our analysis is also affected by another related feature of the dataset. Many large retailers such as Amazon sell, among many other things, music in digital format. Our inability to observe precise consumer behavior within each website therefore prevents us from classifying any visit on websites such as Amazon in a music consumption category. Note, however, that this feature will only lower the number of clicks to music purchasing websites, potentially driving our results in favor of a substitution effect. Last, visits on illegal peer-to-peer file sharing websites do not allow us to differentiate between the file sharing of music files and other types of files such as movies or books. We believe, however, that this variable is still a very good proxy for the ability to obtain recorded music without paying. Note that this feature is likely to overestimate the number of clicks that we identify as illegal music downloading, potentially driving our results in favor of a substitution effect.

3.3.1.2 Variables

Our econometric specification requires the construction of a set of variables that measure or proxy the determinants of legal digital music purchases. Aside from the type of individual socioeconomic characteristics mentioned above, we need variables related to the individual's online activity. First, we expect some other forms of entertainment to be related to the consumption of digital music. For each individual, we therefore consider the number of clicks on websites related to the following activities: online shopping, books & magazines, events, cinema and CDs purchase.[12] Second, we can use information on the visits to specific types of website as a proxy for individuals' taste for music. Individuals with a strong interest in music

[12]Books & magazines websites are sites that contain information, products, and/or services specifically on books and/or magazines. Events websites are sites that contain information and/or tickets sales specifically on physical events. Cinema websites are sites that contain information, products, and/or services specifically on movies, videos, and/or any other products and services associated with the movie industry. CDs purchase websites are sites that allow the purchase of CDs and LPs. These are rather specific websites that sell either collectibles or limited edition CDs. They are not websites from large retailers where one could find any type of CD. The latter type of website is included in the online shopping category. As already mentioned, we are unfortunately not able to identify the visits related to CDs from the ones related to other types of goods on these websites.

are indeed more likely to visit music-related websites such as radio and music-video websites. We also consider websites that are related to music but not to direct music consumption. These include websites related to music news, songs' lyrics or musical instruments. We finally also consider a variable that gives the total time spent online on all the websites of our dataset.

3.3.2 Descriptive Statistics

The following subsection presents some characteristics of the individuals in our sample. We then look at descriptive statistics on the online music consumption behavior for these individuals.

3.3.2.1 Music Users Characteristics

Table 3.1 presents some characteristics of the music users that constitute our final sample. Individuals are, not surprisingly, quite evenly distributed among the 5 different countries and in terms of gender. Almost half of the individuals in the sample have between 31 and 50 years of age, while more than 25% is less than 30 years old.[13] More than 65% of the individuals is employed, with close to 8.5% being self employed, 8.5% are students, and 17% are out of the labor force.[14] The unemployment rate in our sample is of 8.5%. Education level is decomposed into three categories: Primary, secondary and tertiary. Close to 27% of the sample has no more than a primary level of education, and more than a quarter has a secondary level of education. The remaining 47% has a tertiary education level. Total household income is divided into three categories.[15] Twenty-two percent of the sample has a low household income; 62.3% has a medium family income; and the remaining 15.3% has a high household income. Half of the individuals in the sample form part of a less-then-two-people household, while 41% belong to a household of 3 to 4 people. The remaining 8.5% belongs to households of 5 or more individuals.

In terms of music consumption, almost 57% of the individuals have clicked at least once on a legal downloading website. Similarly, 57% of the sample has clicked at least once on a legal streaming website during 2011. Finally, close to 73% of the sample has clicked at least once on an illegal music website during 2011. Note that these different types of music consumers are not mutually exclusive. Figure 3.1 describes the distribution of music consumer types in

[13]The mean age in the sample is 39.7.

[14]These include children under 16, retired, homemakers, full-time carer (of someone in the household) or individuals out of the labor force for other reasons.

[15]For France, Germany, Italy and Spain the income ranges are as follows. Low: Less than 18000 EUR a year. Medium: Between 18000 and 54000 EUR a year. Large: More than 54000 EUR a year. For the UK, the income ranges are as follows. Low: Less than 15000 GBP a year. Medium: Between 15000 and 50000 GBP a year. High: More than 50000 GBP a year.

the sample and reveals that only 40% of the music consumers belong to a single category. Twenty-six percent of the consumers actually belong to the 3 categories. More than half (60%) belong to at least two categories, and 53% of the sample consumes both legal and illegal digital music. Finally, note that 20% of the individuals in the sample have only clicked on illegal downloading websites.

3.3.2.2 Clicks

We now take a closer look at the behavior of the different types of individuals in our data. We can obtain a measure of music consumption intensity by looking at the number of times a consumer clicks on a given website or on a category of specific websites. Table 3.2 presents the mean number of monthly clicks on the different categories of websites (buying, streaming and illegal) as well as the mean number of active months for the individuals in our final sample.[16]

Several interesting patterns emerge when looking at individuals by country. In particular, Spain shows a much larger number of clicks on illegal downloading websites than the other remaining countries, and the second lowest number of monthly clicks on legal music websites. Italy and the UK also show larger number of visits on illegal websites, but Italy presents the lowest number of visits on legal webpages. Gender differences are also important in terms of illegal clicks, but not so much for legal (purchase and streaming) websites. Males show a much larger number of monthly clicks on illegal music websites. In terms of age, individuals between 16 and 40 also have an above average number of monthly visits on illegal music websites, with a rather low number of visits on legal pages. The same observation holds for students (and to a lesser extent for unemployed) when compared to individuals with other employment status. The mean number of clicks on illegal downloading websites is substantially lower for higher income categories.

The figures show that legal consumers (individuals that never clicked on an illegal music website during 2011) are, on average, active 2.5 months a year, while downloaders are active almost 6 months a year. Most interestingly, downloaders are also more active than legals both in terms of legal downloading (10% more clicks) and legal streaming (40% more clicks), as shown by their mean values of clicks. A positive relationship between legal and illegal consumption of digital music therefore emerges from this simple comparison of means. Comparing streamers and non-streamers (individuals that never clicked on a streaming music website during 2011) leads to similar conclusions. The figures show that streamers click more than twice as much on legal downloading websites, while their clicks on illegal downloading websites is 90% higher than for non-streamers. Again, this simple comparison of

[16] An active month is defined as a month in which the individual visited at least one of the three categories. Note that this definition does not take into account the intensity of clicks within an active month.

means shows a positive relationship between the different consumption channels and in particular between streaming and legal purchases. Table 3.3 presents cross correlations between the different music consumption channels and gives further evidence on this point. All these figures suggest, not surprisingly, that music taste is an important determinant of digital music consumption, regardless of its origin. In other words, one should expect people who like music to consume more of it, whether it is through legal downloading, illegal downloading, or streaming.

3.4 Research Question and Empirical Approach

The impact of music piracy on legal sales of music has been studied extensively in the empirical literature, focusing mainly on legal music sales in the form of physical CDs. While most studies find that piracy harms revenues, the estimated sales displacement rate is far below one. In other words, music consumers are found to substitute legal music consumption for illegal music consumption, but much of what is consumed illegally would not have been purchased if piracy was not available. Since the launch of the iTunes music store in 2003, the availability to purchase legal digital songs changed individuals' music consumption alternatives. Instead of having to buy a whole CD, the alternative to downloading any particular digital song illegally is now to purchase it in MP3 format. As emphasized in Waldfogel (2010), the appearance of file-sharing and downloading technology might have different effects on sales, depending on whether the legal option is a 12-song CD or à la carte songs. Consider an individual interested in a few songs of a given artist. While she may not consider buying the entire album (which also contains *unknown* songs) when offered the possibility to freely download these specific songs, she might nevertheless be willing to pay for them individually. The effect of downloading on individual songs and albums may therefore be different, and one can easily imagine a circumstance in which file-sharing would hurt album sales more than it hurts song sales.

Our goal is to answer two broad questions. First, we are interested in looking at the determinants of music consumption in the form of purchasing, downloading and streaming. The motivation for this descriptive exercise is to understand better the demographic characteristics that drive music consumption through the different channels available to consumers. Our second objective is to see to what extent these different channels are related to each other. We revisit the sales displacement question and ask whether illegal music downloading is used as a substitute for legal digital music consumption. We are also interested in the effect that legal streaming may have on legal digital music consumption. As in the case of file-sharing, economic theory does not provide us with an unambiguous prediction for how music streaming should affect purchases of digital music. On the one hand, consumers may substitute legal downloads for streaming. On the other hand, consumers may well use streaming to sample new artists and/or songs. In particular, it may be the case that individuals assign a higher value to a song when they posses it, as opposed to simply having access to it. This would enhance the value of streaming services as discovering tools, which would positively affect sales. Another part of the debate on streaming media is related to the concept of "windowing", a strategy used by some artists requesting that an album be available first only on sales before being available on streaming platforms.Understanding whether consumers use these two channels as complements or substitutes is therefore crucial to understand how these types of strategies actually affect sales.

3.4.1 Determinants of Music Consumption

Before starting to analyze the effects of illegal downloading and streaming on legal music purchases, we propose to look at the determinants of these three different music consumption channels.

The empirical literature on the determinants of digital music consumption is rather thin. In particular, very few studies have analyzed the factors that influence the legal consumption of digital music such as online purchases or online streaming. However, most of the papers that analyze sales displacement of digital music also provide some evidence on the factors influencing these purchases. In general, age does not seem to affect purchasing behavior and no significant nor systematic difference is found between males and females. Income is, however, positively correlated with online sales of digital music (Bastard et al., 2012; Cecere et al., 2012; DangNguyen et al., 2012; Waldfogel, 2010). Probably unsurprisingly, all studies show that legal consumption of online music is increasing in the individuals' interest for music.

Studies on the determinants of illegal music consumption are also rather limited, although papers analyzing music sales displacement again provide some evidence on factors affecting online piracy (Zentner, 2006; Waldfogel, 2010). Clearer patterns emerge from this literature. Gender seems to matter in terms of music piracy behavior, with males being much more active. In a large sample of European consumers, Zentner (2006) finds that online music piracy is negatively correlated with both education and income, although not significantly so for the latter. Again unsurprisingly, all studies show that there is a strong correlation between illegal music consumption and interest in music.

Although the existing empirical literature on the factors influencing digital music piracy is scarce, there is a rather important body of literature on the determinants of digital piracy of other sorts of products such as software. Looking at the patterns found in these studies may therefore be informative to further investigate the determinants of online music piracy.

According to this literature, some of the most relevant factors influencing piracy are income, cultural differences, past behavior or habit formation, and the legal setting.[17] The vast majority of studies finds that income is negatively correlated with digital piracy. Culture plays an important role, too. In a study analyzing software piracy, Marron and Steel (2000) find that countries characterized by individualist cultures have higher piracy rates than countries with a collectivist culture. They also find that piracy rates are lower in countries that have strong institutions that enforce contracts and protect property from expropriation. Another set of studies uses individual data to study the factors that determine piracy behavior.[18] They also confirm that piracy is negatively correlated with income and show that it is generally higher for male than for females. Most of these studies are, just as in the case of the above-mentioned music studies, based on college students' surveys. As such they do not allow for a clear analysis of the relationship between education levels and piracy. Likewise, they provide only limited evidence on the correlation between age and piracy given that the range of ages is quite limited within such population. Studies analyzing this specific question are therefore rather scarce. One exception is the work of Mandel and Süssmuth (2012) who study a sample representative of the German working population with high-speed internet access.[19] They, too, find that frequency of digital piracy is negatively correlated with income. Although they find no significant gender difference in the propensity to pirate, they find that male individuals are prone to pirate at a significantly larger scale. Finally, their findings indicate that individuals in their early twenties (between 20 and 25 years old) are predominantly responsible for the overall extent of digital piracy in their sample.

Using the cross-sectional dimension of our data, we turn to the analysis of the determinants of music consumption through the different channels available to consumers. Our objective is to describe how the number of clicks on purchasing, downloading and streaming websites vary across individuals in our sample. Table 3.4 presents the results of this exercise. Each column of the table represents the regression of the different dependent variables (the clicks on purchasing, downloading or streaming websites) on the same set of regressors.

Considering first demographic characteristics, some differences are worth noticing. Legal purchases of digital music raise with household income and seem to be more prevalent among males. Education, on the other hand, seems to have no significant effect on the legal purchases of digital music. In line with the results of the previous literature presented above, there is a strong negative correlation between income level and illegal downloading activity. As in the case of legal purchases, digital music piracy seems to be a predominantly masculine activity

[17] See Waldman (2013) for a review of the literature analyzing the different factors determining piracy of intellectual property.

[18] See for example Ramayah et al. (1990), Sims et al. (1996), Limayem et al. (2004) and Ramayah et al. (2009).

[19] Although they integrate it in their piracy measures, the authors do not confine their analysis to digital music piracy.

and no clear pattern emerges on its relationship with education levels. The determinants of music consumption via online streaming services show different patterns. There seem to be no significant gender nor income differences in this specific mode of music consumption. Education is, however, positively correlated with online streaming of music. Finally, music consumption appears to differ significantly by age groups only when it is in the form of illegal downloading or online streaming. Individuals aged between 21 and 25 stand out as being the most active in terms of illegal downloading while online streaming seems to be an activity of the really young.

The country differences are remarkable for the three modes of music consumption. In terms of purchases, Spaniards and Italians have 50% less clicks than Germans, British have 22% less clicks and French 13% less. These differences could relate to several characteristics captured by our country dummies. First of all, it might be that not all of the countries have the same availability of legal digital purchasing websites, and purchasing a song from another country's website is not always feasible. The same considerations hold when thinking about online streaming services. For that specific type of music consumption, France stands out with 150% more clicks than Germany. Spaniards have 20% more clicks than the German, while Italians have 25% less. The UK presents a small difference with Germany in terms of streaming, with only 9% more clicks. These differences are again possibly due to differences in availablity, especially given that online streaming services were a rather new music consumption mode in European countries in 2011. For example, neither Italy nor Germany had access to the Spotify online streaming service as of 2011. Individuals' awareness of the existence of such services may therefore not be equal in all countries, affecting their ultimate usage levels. The most striking differences appear when looking at the determinants of illegal downloading. Used as the reference country, Germany shows the lowest level of visits on illegal downloading websites. Compared to it, Spain and Italy show important differences of 230% and 134% more clicks respectively. Individuals from the UK have 43% more visits while French individuals present a differential of 35%. As mentioned above, several non-mutually exclusive explanations could drive these important country differences. Again, market forces, and in particular the limited access to legal digital purchasing websites, could influence the illegal downloading activity of consumers.[20] Second, unobservable cultural characteristics could explain the use of different types of music consumption channels. In particular, past behavior and cultural factors are, as highlighted above, important determinants of digital piracy. Individuals from different countries may also differ in their cultural norms or standards toward acceptable behavior and may perceive differently the extent to which their piracy behavior affects artists and/or producers. It may also be the case that indivuduals more used to downloading illegaly (say because legal sources to obtain digital music were not

[20] For the case of television content, Danaher et al. (2010) present evidence that the lack of legal channels can positively affect the level of piracy.

previously available) may stick to this illegal behavior as a consequence of habit formation. Finally, previous literature showed how different legal setting influence piracy behaviour. Cross-country differences in individual piracy behavior may indeed also be driven by differences in specific, national copyright enforcement laws (e.g. the HADOPI law in France). The important significance of our country variables indicate that all of the above mentioned factors play an important role in the determinants of online music consumption.

The three types of music consumption are positively and significantly increasing in the variables that capture interest in music (visits on music related websites). This unsurprisingly confirms that individuals who like music enjoy consuming more of it via the different channels available. The coefficients on the variables related to other online activities present some differences as well. The visits on book websites are positively correlated with purchasing and streaming, but not with downloading. Clicks on events websites are positively correlated with purchasing and streaming, but negatively with downloading and movies websites are positively correlated with all three channels of music consumption. This is also true for visits on types of websites related to instant messaging and personal webpages. Finally, clicks on global news and social network webpages significantly affect downloading and streaming, but not purchasing of digital music.

3.4.2 Displacement: Downloading, Streaming and Purchases

We now turn to the effect of illegal downloading and streaming on legal music purchases. In particular, the question we want to answer is how much does an instance of downloading (respectively streaming) depress or stimulate digital music purchases. Ideally, we would like to compare the legal purchases of an individual who has access to downloading (streaming) with the legal purchases of that same individual in the hypothetical case in which she has no access to downloading (streaming). This direct comparison is obviously impossible, as no individual can simultaneously be in these two scenarios. Since we only observe consumers when they have access to downloading and streaming, we have no way of knowing directly how they would have behaved had they had no access to those services.

One can start by asking whether individuals who download (stream) more also purchase more. The correlations already presented in table 3.3 showed the positive relationship between the different music consumption modes. The main problem of this simple approach is that individuals who like music like to consume more of it through the various channels available. This would give rise to a positive relationship between downloading (respectively streaming) and digital music purchases, regardless of whether a complementarity relationship exists. This prevents us from giving a causal interpretation to this positive relationship, as we have no way of knowing how an exogenous change in the availability of illegal downloading (respectively streaming) would affect legal purchases. The main obstacle therefore comes from individual unobserved characteristics, and in particular their taste in music. Several approaches can be used to circumvent this problem. One is to look for some measures of interest in music in order to partially control for unobserved individual heterogeneity. We use information on online behavior by considering the number of clicks on music-related websites such as radio and music-video websites. We also consider sites that are related to music, although not to direct music consumption. These include websites related to songs' lyrics, musical instruments or music news such as blogs. Note that, contrary to the indicators used in previous studies, these variables have several advantages. First, they are not the result of a subjective assessment from the individual. In many survey-based studies, music taste is measured by asking individuals about their music taste on a numerical scale (Rob and Waldfogel, 2006; Zentner, 2006; Waldfogel, 2010). Such a measure is plagued with several problems. Different people will assign different meanings to it (a *strong taste* in music may not have the same meaning for individual A than for individual B), making it an imperfect indicator of music interest. Also, category-based variables are less informative than variables that actually measure the strength of the factor of interest. Our measure of music taste avoids this types of problems. First, no self-assessment from the individual is needed as it is the result of directly observed behavior. Second, our data not only allows us to observe whether an individual visited a given music-related website, it also gives us a measure of the number of times such visit was made. This gives us a better measure of the intensity of the factor we want to capture, namely the interest in music. We therefore believe our variables to be more reliable indicators of music interest than standard survey-based measures.

3.4.2.1 Cross-sectional approach

We start by looking at cross-sectional regressions of the following form:

$$P_i = X_i\beta + W_i\alpha + \delta D_i + \gamma S_i + \varepsilon_i, \tag{3.1}$$

where for individual i, P_i is the (log of the) number of clicks on legal purchase websites, D_i is the (log of the) number of clicks on illegal downloading websites, S_i is the (log of the) number of clicks on legal streaming websites, X_i is a vector including socioeconomic characteristics of the individual, and W_i includes a set of variables related to the individual's online activity on other types of websites. Unobserved characteristics affecting individual i's clicks on legal purchase websites are included in ε_i, and α, β, γ and δ are parameters to be estimated. The unobserved heterogeneity problem in this specification comes from the fact that we expect ε_i to be correlated with D_i and S_i due to unobserved taste in music. Our measures of music interest in the form of visits to music-related websites is therefore included in W_i to solve that problem. The identifying assumption is therefore that, once controlling for music taste and other observable characteristics, the number of clicks on downloading and streaming websites is random.

Table 3.5 reports estimates of equation (3.1) using OLS. In this equation the unit of observation is an individual and X includes country dummies and individual socioeconomic characteristics. The specification in column 1 only controls for X and reports significant and positive estimated coefficients for both γ and δ. Columns 2 includes our main measure of music interest, the number of visits to music-related websites. Controlling for this variable leads in a 20% drop in the effect of both downloading and streaming on purchases. Columns 3 to 5 include other measures of music consumption such as radio, illegal streaming and specific CD purchases.[21] As expected, the introduction of such variables decreases the estimates of γ and δ. In particular, comparing columns 1 and 5, the estimated effect of clicks on illegal webpages drops by close to 40% when introducing the variable measuring the visits on other music websites. In turn, the estimated effect of the time spent on streaming websites drops by 35% . In column 6 we include more explanatory variables related to other forms of entertainment websites. All show positive and significant effects, except for websites related to movies and personal webpages. When we include the complete set of regressors (column 6), our coefficients of interest remain positive and highly significant. The estimates reveal positive elasticities of about 0.03 and 0.08 for the illegal music downloading websites and legal streaming websites respectively. Our results also show interesting country differences in terms of legal purchase of legal digital music. Individuals from Spain and Italy show around 50% less clicks than Germany, while the UK and France present around 20% less of such clicks.

Given that visits to legal music purchases websites is equal to zero for over fifty percent of the observations, we estimate again equation (3.1) using a Tobit model. Table 3.6 reports the unconditional marginal effects of the estimation. The estimate for δ drops by half while the estimate for γ diminishes in one third.

Our results suggest that illegal downloading and legal streaming have both a positive and significant effect on legal purchases of digital music. Although we have constructed measures of individuals' interest in music using online activity measures, we cannot completely rule out the existence of other forms of unobserved heterogeneity we are not able to control for. While visits to different kinds of music related websites surely capture individuals' interests in music, individuals who do not visit these specific types of webpages may still differ substantially in their taste for music. In other words, although our online measures allow us to control

[21]CDs purchase websites are sites that allow the purchase of CDs and LPs. These are rather specific websites that sell either collectibles or limited edition CDs. They are not websites from large retailers where one could find any type of CD. The latter type of website is included in the online shopping category. As already mentioned, we are unfortunately not able to identify the visits related to CDs from the ones related to other types of goods on these websites.

for an important part of individuals' music taste, they may still not allow us to capture it in its entirety, especially if individuals do not reveal their music taste through their online activity. For instance, consider individuals A and B who differ only in that B has no interest in music. It may well be that neither A nor B visit any of the other music-related websites. In that case our measure would not take into account the fact that individuals A and B still differ drastically in their taste for music. Since we expect our results to be biased away from finding a negative effect of downloading (respectively streaming) on purchases, finding a positive result may simply reflect the fact that our estimations are still contaminated by individual unobserved characteristics. We next turn to an empirical strategy that allows us to tackle this problem.

3.4.2.2 Longitudinal approach

The second approach that we use to further solve the problem of unobserved heterogeneity consists in exploiting the panel dimension of our data. Cross section regressions estimate the effect of illegal downloading (legal streaming) on legal purchases by comparing individuals with low levels of illegal downloading (legal streaming) and high levels of illegal downloading (legal streaming). The panel structure of the data allows us to take advantage of the variation in these variables within individuals and to control for other (time invariant) unobservable individual determinants of music consumption. We first consider the following regression equation

$$P_{it} = X_i\beta + W_{it}\alpha + \delta D_{it} + \gamma S_{it} + \xi_t + \varepsilon_{it}, \tag{3.2}$$

where the unit of observation is now an individual per month. Thus we regress the number of clicks on legal purchase websites made by individual i in each month on the number of that month's clicks on illegal downloading websites and legal streaming websites along with monthly time dummies ξ_t and our previous controls. Estimating (3.2) by pooled OLS allows us to take advantage of the within individual variation in P_{it}, D_{it} and S_{it} when estimating δ and γ.

Table 3.7 present the results of estimating (3.2) by pooled OLS using data for all individuals in all months. We cluster standard errors at the individual level since the error term ε_{it} is likely to be correlated over time for a given individual. The estimates of δ and γ are reduced in about 15% compared to the cross sectional estimations. The estimates suggest elasticities of about 0.025 and 0.07 for the illegal music downloading websites and legal streaming websites respectively. The coefficients on the monthly time dummies present evidence of some seasonal effects. Taking July as a reference, it seems that visits on legal purchase websites are higher from December to March, and lower from August to November. These differences could be driven by several factors. For instance, a larger number of clicks on legal purchasing websites during the months following December could be the result of gift cards bought and exchanged during the Christmas season. Likewise, the lower number of clicks observed during the summer could be related to holidays when individuals may not be spending as much time on their home computer.

No significant differences are noticed from April to June. The coefficients on the regressors related to online activity show estimates that are lower in magnitude as compared to the cross-sectional estimation.

We again estimate equation (3.2) using a Tobit model in order to take into account the fact that monthly visits to legal music purchases websites are 0 in our data. Table 3.8 reports the unconditional marginal effects of the estimation. The estimate for δ drops by half and the estimate for γ diminishes in two thirds. The elasticity of downloading clicks is therefore similar to the one found when using only cross-sectional variation, while the elasticity of streaming clicks is somewhat lower.

Our estimates in tables 3.1-3.8 might still be vulnerable to the concern that illegal downloading and streaming are endogenous. As mentioned above, it may well be that some other form of unobserved heterogeneity is not completely captured in our measures of music interest. While people who visit many music-related websites have most certainly a high interest in music, it may be that some individuals with a high taste for music don't visit such webpages often and only click on websites that allow them to download, stream or purchase songs. The longitudinal structure of our data allows us to deal with this concern and to further control for fixed unobservable individual characteristics. We make the substitution $\varepsilon_{it} = \mu_i + \nu_{it}$, where μ_i is an individual-specific fixed effect and ν_{it} is an individual and month-specific error, and estimate the following equation

$$P_{it} = X_i\beta + W_{it}\alpha + \delta D_{it} + \gamma S_{it} + \xi_t + \mu_i + \nu_{it}. \tag{3.3}$$

Fixed-effects estimation allows to control for time invariant unobserved heterogeneity (such as interest in music) and identifies coefficient δ (respectively γ) from the relationship between variation in the tendency to click on legal purchase websites and variation in the tendency to click on illegal downloading websites (respectively streaming websites) for each individual. Only within individual variation is therefore used to identify our parameters of interest. Note that this estimation strategy allows us to control for both time invariant taste in music (captured in the individual fixed effects) and possibly time variant shocks to music taste that are captured by the visits on the music-related websites.

Table 3.9 presents the results of the estimation of equation (3.3). Including individual fixed effects we obtain coefficient estimates of 0.022 for our illegal downloading variable. In other words, a 10% increase in clicks on illegal downloading websites is associated with a 0.2% increase in clicks on legal purchasing websites. This effect is larger for legal streaming, with a 10% increase in clicks on these websites being associated with a 0.49% increase in clicks on legal purchasing websites. The reduction in both these coefficients as compared to the ones that correspond to our pooled OLS estimation (table 3.7) therefore show the importance of taking individual fixed effects into account.

3.4.2.3 Country specific effects

The results presented in table 3.4 describe remarkable cross-country differences in the individuals' number of clicks on each of the three alternative music consumption channels. As discussed above, several non-mutually exclusive explanations related to country-specific characteristics can drive these differences such as cultural traits, market forces or the enforcement and effectiveness of copyright-specific laws. All of these effects are captured by our country-specific dummy variables and we are unfortunately not able to identify the relative importance of these different mechanisms. Nevertheless, it is natural to consider that these country differences may also influence the displacement rates of legal purchases of digital music by illegal downloading and legal streaming. For instance, individuals with different perceptions of piracy are likely to have different music consumption habits. In particular, consumers with more permissive attitudes toward piracy are probably more likely to substitute legal consumption of digital music by illegal consumption and should therefore present higher displacement rates. Likewise, individuals coming from countries with more stringent copyright laws will be affected by the latter when deciding on consuming pirated content and may refrain from using this type of consumption channel to a larger extent. To check for the possibility that underlying country-specific characteristics could lead to different displacement rates, we expand equation (3.3) and estimate the following specification:

$$P_{it} = X_i\beta + W_{it}\alpha + \delta D_{it} + \sum_{c\in C}\delta_c D_{it} Country_{ic} + \gamma S_{it} + \sum_{c\in C}\gamma_c S_{it} Country_{ic} + \xi_t + \mu_i + \nu_{it}, \quad (3.4)$$

where $Country_{ic}$ is a dummy variable equal to 1 if individual i is from country $c \in C$ and $C = \{$ Spain, Italy, France, UK $\}$. The parameter δ_c (γ_c) measures the difference between the effect of downloading (streaming) on purchases in country c compared to the effect of the same variable in Germany.

Table 3.10 present the results of estimating equation (3.4) using fixed effects (within) estimation. The results show no evidence of sales displacement for any of our countries. Note how each country specific coefficient decreases for both illegal downloading and legal streaming (i.e. the sums $\delta + \delta_c$ and $\gamma + \gamma c$, respectively) as we include our explanatory proxy variables for music interest, confirming again the importance of controlling for these factors. Focusing on the relationship between legal streaming and legal purchases of digital music, results show statistically significant cross-country differences. French users present the largest coefficients (about 0.068) followed by the Germans (0.061) and by UK users (0.044). Users from Spain and Italy present elasticities of 0.022 and 0.033, respectively.[22] Interestingly enough, there seems to be no clear relationship between the intensity of clicks on legal streaming websites as presented in table 3.4 and the elasticities of legal streaming. For instance, while users from Spain have a much higher intensity of clicks on legal streaming websites compared to Italian users, they show statistically indistinguishable coefficients in table 3.10.

Although there is no displacement of legal purchases by illegal downloading to be found, the cross country differences are, again, remarkable. Looking at the relationship between illegal downloading and legal purchases, French users again present the largest coefficients (about 0.044) followed by the Germans (0.038) and by UK users (0.026).[23] The most striking difference appears for Spain and Italy which show very small elasticities of 0.008 and 0.009, respectively. This means that a 10% increase in the clicks on illegal downloading websites is associated with almost no change in visits to legal purchasing websites for Spanish and Italian users (0.08% and 0.09% increases, respectively). As already highlighted above, our country dummy variables capture any unobservable country-specific characteristics that would affect individuals' behavior toward music consumption. We can nonetheless try to relate some of the important country differences that we found in the determinants of digital music consumption to the elasticities displayed in table 3.10. Speculating beyond what our analysis allows us to show, the significant differences in elasticities between illegal downloading and legal purchases could for instance be the result of differences in the availability of legal digital music stores. In 2011, neither Spain nor Italy had seen the entry of the Amazon mp3 online music store, as opposed to France, Germany and the UK.[24] This relative lack of legal online outlets could therefore have made consumers more exposed to piracy for a longer period of time given its

[22] The difference between users from Germany and France are not statistically significant and neither is the difference between users from Germany and the UK. The difference between users from France and the UK are, however, statistically significant.

[23] As in the case of the coefficients on the legal streaming variable, the difference between users from Germany and France are not statistically significant and neither is the difference between users from Germany and the UK. The difference between users from France and the UK are, however, statistically significant.

[24] The Spanish and Italian editions of the Amazon mp3 store were launched on Octobre 4, 2012. it was launched on December 3, 2008 in the UK, on April 1, 2009 in Germany and on June 10, 2009 in France. See `http://en.wikipedia.org/wiki/Amazon_MP3`.

relative appeal.[25] Given the important link between habit formation and attitude toward piracy found in earlier research, consumers that are more experienced with digital music piracy could build more permissive views on it and therefore show higher displacement rates (i.e. lower elasticities).[26]

3.4.2.4 Displacement and interest in music

We have seen how controlling for interest in music and individual unobserved heterogeneity is crucial to determine the relationship between the different online music consumption channels. Taking such factors into account is key because individuals with higher interest in music are likely to consume more music through the different channels, which would give rise to a positive correlation regardless of whether a complementary relationship exists. If we expect consumers with different degrees of interest in music to make different consumption decisions, it seems natural to ask whether different levels of music interest are associated with different elasticities between the alternative music consumption channels. For instance, we may expect individuals with higher levels of music interest to show lower displacement rates (i.e. larger elasticities). One may indeed argue that these individuals might indeed use illegal downloading more as a discovery tool.[27] Such users may also be more likely to use online streaming services as discovery tools, which would lead to higher elasticities of legal online streaming as well. In our setup, a higher taste for music is measured by a larger number of visits on music related websites (such as radio and music-video websites) and to other types of websites related to music, although not necessarily to direct music consumption (such as websites related to songs' lyrics or music news). To check whether individuals with higher levels of music interest present different displacement rates, we extent equation (3.3) by interacting our music interest variables with the visits to illegal downloading websites and legal streaming websites. Table 3.11 presents the results of this exercise. The first column shows the results of estimating equation (3.3) without any interaction terms and actually corresponds to column (6) in table 3.9. For variables measuring visits to illegal downloading websites and legal online streaming websites, specification (2) introduces interaction terms with visits on radio and music video websites. Specification (3) introduces interactions with visits to other types of music websites (although not directly related to music consumption). Finally, column (4) incorporates all interactions used in specifications (2) and (3). Our estimates show that there are indeed significant elasticity differences as a function of the intensity of visits to other types of music websites. For the visits on illegal downloading

[25] It is important to note that we are not able to explain why the level of piracy is higher in certain countries. In particular, the higher levels of piracy observed in Spain and Italy could be higher than in the other countries of our sample for other reasons than the availability of legal online music stores.

[26] Most of the literature on the link between habit formation and attitude toward piracy focuses on software piracy. See for instance Limayem et al. (2004) and Ramayah et al. (2009).

[27] Likewise, individuals with higher interest in music may put a higher value on music as a cultural good and have less permissive views on piracy.

websites, the last 3 columns of table 3.11 show that individuals with higher visits to other types of music websites (i.e. individuals with higher interest in music) have much higher elasticities. Using the estimates from specification (3), a 10% increase in the number of clicks on illegal downloading websites is associated with a 0.18% increase in clicks on legal purchasing websites for an individual with a log number of clicks on illegal downloading websites equal to the sample mean. For values of the log of clicks on illegal downloading websites at the 90th and 95th percentiles of the sample, this effect increases to 0.33% and 0.4%, respectively. Interestingly, this effect is unchanged by the inclusion of the interaction with visits to radio and music video websites (see column (4)). Similarly, our results show that a higher interest in music is associated with larger elasticities of visits to legal streaming websites. It seems, however, that the effect of visits to radio and music video websites is larger for the elasticity of legal streaming than it is for the elasticity of illegal downloading. Using the estimates from specification (4), and evaluating the effects at the sample mean, we find that a 10% increase in the number of clicks on legal streaming websites is associated with a 0.3% increase in clicks on legal purchasing websites. When evaluated at the 90th and 95th sample percentiles, this effect increases sharply to 0.58% and 0.71%, respectively. Our results therefore show that individuals with a larger interest in music (in the sense of visiting music related websites to a larger extent) seem to use online legal streaming as a discovery tool that further stimulates their visits to legal purchasing websites.

3.5 Conclusion and Discussion

In the last decade, the music industry has faced many changes. In particular, it has seen its revenues decrease drastically, with industry representatives blaming most of it on piracy (IFPI, 2011). Nevertheless, the music industry seems to have embraced digitization and its many business opportunities. Indeed, digital music revenues have increased more than 1000% during the period 2004-2010, growing 8% globally in 2011 to an estimated US$5.2 billion (IFPI, 2011, 2012). While most empirical studies have indeed confirmed a significant negative impact of piracy on sales of physical music, the growing importance of the digital sector in total music industry revenue calls for a better understanding of the impact of both piracy and other music consumption channels on legal digital sales. In this paper, we revisit the question of music sales displacement in the digital era, and analyze in detail the effect of online music streaming on the legal purchases of digital music.

Conducting research on the revenue effects of illegal music consumption requires detailed data on the quantities of both legal and illegal music consumed by individuals. Relying on an original dataset, we are able to follow the clickstreams of more than 16,000 Internet users, and in particular their visits to legal and illegal music consumption websites.

After using several approaches to deal with the endogeneity of downloading and streaming, our results show no evidence of online digital sales displacement. Overall, our different estimates show relatively stable, positive and low elasticities of legal purchases with respect to both illegal downloading and legal streaming. Across specifications, the estimates of δ suggest elasticities of about 0.02 between clicks on illegal downloading websites and legal purchases websites. If this estimate is given a causal interpretation, it means that clicks on legal purchase websites would have been 2% lower in the absence of illegal downloading websites. Specific country estimate show that for Spain and Italy the elasticity is zero, while it is close to 0.04 for France and Germany and close to 0.03 for the UK. All of these results suggest that the vast majority of the music that is consumed illegally by the individuals in our sample would not have been legally purchased if illegal downloading websites were not available to them.

The existing literature analyzing the effect of piracy on digital music consumption is still limited. Our study contributes to this literature with results in line with the findings of Bastard et al. (2012) which show that illegal music downloads have little or no effect on legal digital sales. It is very important to understand that these results need not contradict earlier research that found substantial amounts of sales displacement of legal physical music sales by illegal digital downloads. One must realize that music consumption in physical and digital format are two very different consumption modes, and it is easy to imagine circumstances in which music piracy would affect sales of albums in CD format differently than it would affect sales of individual digital songs. Considering an individual interested in a specific song, the trade-off between a whole 12-songs album and a free pirated version of that song is rather different from the trade-off involving the legal purchase of the individual digital song alone. As such, sales displacement from piracy in a world where only CDs are available does not necessarily imply sales displacement in a world where digital à la carte songs are made available to consumers.

Another contribution of our paper is the analysis of the effect of online music streaming on the legal purchases of digital music, a question that has received very little attention in the empirical literature thus far. On this particular question, our elasticity estimates show somewhat larger figures, ranging from 0.024 in our Tobit specification to 0.07 in the OLS case. Controlling for individual fixed effects leads to a 0.05 elasticity, suggesting complementarity between streaming services and purchases of legal digital music. Again, country differences show that this effect is larger for France and the UK (around 0.06) while it is smaller for Spain and Italy (around 0.035). Our results are in line with the results in (DangNguyen et al., 2012), the only study that has, to our knowledge, analyzed the question so far.

Taken at face value, our findings indicate that digital music piracy does not displace legal music purchases in digital format. This means that although there is trespassing of private property rights (copyrights), there is unlikely to be much harm done on digital music revenues. This result, however, must be interpreted in the context of a still evolving music industry. It is in particular important to note that music consumption in physical format has until recently accounted for the lion's share of total music revenues. If piracy leads to substantial sales displacement of music in physical format, then its effect on the overall music industry revenues may well still be negative.

We cannot draw policy implications at the industry-wide level, as our analysis is only confined to the digital segment of the music industry. Nonetheless, digital music revenues to record companies are growing substantially, reflecting the increasing importance of digitization in the music industry (IFPI, 2012). From that perspective, our findings suggest that digital music piracy should not be viewed as a growing concern for copyright holders in the digital era. In addition, our results indicate that new music consumption channels such as online streaming positively affect copyrights owners.

Table 3.1: Individual characteristics: music users†

	No. of individuals	%	Cumul. %
Country			
France	3386	20.8	20.8
Germany	3091	19.0	39.8
Italy	3281	20.1	59.9
Spain	3664	22.5	82.4
UK	2868	17.6	100.0
Total	16290	100.0	
Gender			
Female	7892	48.4	48.4
Male	8398	51.6	100.0
Total	16290	100.0	
Age Category			
10-15	692	4.2	4.2
16-25	2062	12.7	16.9
26-30	1657	10.2	27.1
31-40	4278	26.3	53.3
41-50	3911	24.0	77.3
51-60	2338	14.4	91.7
61-75	1352	8.3	100.0
Total	16290	100.0	
Employment			
Employed	9371	57.5	57.5
Out of Labor Force	2775	17.0	74.6
Self Employed	1375	8.4	83.0
Student	1388	8.5	91.5
Unemployed	1381	8.5	100.0
Total	16290	100.0	
Education			
Primary	4359	26.8	26.8
Secondary	4233	26.0	52.7
Terciary	7698	47.3	100.0
Total	16290	100.0	
Household Income			
Low	3649	22.4	22.4
Medium	10144	62.3	84.7
High	2497	15.3	100.0
Total	16290	100.0	
Household size			
1 - 2	8235	50.6	50.6

Continued on next page

Table 3.1: Individual characteristics: music users†

	No. of individuals	%	Cumul. %
3 - 4	6662	40.9	91.4
5+	1393	8.6	100.0
Total	16290	100.0	
Buyer			
No	7070	43.4	43.4
Yes	9220	56.6	100.0
Total	16290	100.0	
Streamer			
No	6978	42.8	42.8
Yes	9312	57.2	100.0
Total	16290	100.0	
Downloader			
No	4457	27.4	27.4
Yes	11833	72.6	100.0
Total	16290	100.0	

† The sample includes all music users, i.e. individuals that either buy, stream or download. Buyers are defined as individuals that clicked on at least one legal downloading website during 2011. Streamers are defined as individuals that clicked on at least one legal streaming website during 2011. Downloaders are defined as individuals that clicked on at least one illegal music website during 2011.

Table 3.2: Monthly Click Activity

	Mean					**N**
	Active Months	Buying	Streaming	Downloading	All	N
Country						
France	4.81	1.65	3.40	6.49	11.54	40,632
Germany	4.28	1.79	1.33	6.24	9.39	37,092
Italy	4.69	0.37	0.98	7.97	9.35	39,372
Spain	5.97	0.41	2.12	10.38	13.11	43,968
UK	4.46	1.23	2.51	7.99	11.75	34,416
Gender						
Female	4.51	1.02	2.06	5.88	9.00	94,704
Male	5.24	1.12	2.09	9.76	13.03	100,776
Age						
10-15	3.93	0.65	1.82	3.82	6.30	8,304
16-25	6.00	0.84	3.41	10.46	14.79	24,744
26-30	5.78	1.61	2.86	10.66	15.21	19,884
31-40	5.18	1.04	2.09	8.53	11.72	51,336
41-50	4.66	1.21	1.91	7.40	10.57	46,932
51-60	4.15	0.86	1.29	6.83	9.03	28,056
61-75	3.55	1.01	1.01	3.76	5.79	16,224
Employment						
Employed	4.92	1.15	2.03	8.16	11.41	112,452
Out of Labor Force	4.04	0.89	1.71	5.22	7.84	33,300
Self Employed	4.68	1.01	1.24	6.70	9.02	16,500
Student	5.97	0.76	3.09	10.19	14.09	16,656
Unemployed	5.43	1.22	2.93	10.14	14.37	16,572
Education						
Primary	4.61	1.40	2.00	7.28	10.73	52,308
Secondary	5.06	0.76	1.91	8.89	11.64	50,796
Terciary	4.95	1.05	2.21	7.66	10.97	92,376
Household Income						
Low	5.34	1.24	2.64	9.70	13.68	43,788
Medium	4.88	1.02	1.81	7.81	10.69	121,728
High	4.25	1.01	2.33	5.51	8.87	29,964
Household size						
1 - 2	5.00	1.29	2.18	8.40	11.92	98,820
3 - 4	4.80	0.84	1.82	7.39	10.12	79,944
5+	4.61	0.84	2.66	7.16	10.68	16,716
Children at home						
No	5.01	1.16	2.25	8.39	11.86	128,556
Yes	4.65	0.89	1.74	6.89	9.58	66,924
Legal	2.49	0.99	1.58	-	2.58	53,484
Downloader	5.79	1.10	2.26	10.85	14.28	141,996
Non-Streamer	3.51	0.66	-	5.20	5.88	83,736
Streamer	5.92	1.37	3.63	9.89	14.98	111,744
Total	4.88	1.07	2.07	7.88	11.08	195,480

† Buying, Streaming and Downloading clicks are defined as clicks on a legal downloading, streaming and illegal downloading websites, respectively. Streamers are defined as individuals that clicked on at least one legal streaming music website during 2011. Non streamers are defined as individuals that never clicked on legal streaming music website. Downloaders are defined as individuals that clicked on at least one illegal downloading music website during 2011. Legals are defined as individuals that never clicked on an illegal music websites. The figures in the table represent the mean number of monthly clicks.

Table 3.3: Cross-correlations of number of clicks

Variables	Buying	Downloading
Buying	1	
Downloading	0.0559*	1
Streaming	0.3634*	0.0470*

* Significant at the 1% level.

TABLE 3.4: Determinants of music consumption, †

	Purchase		Downloading		Streaming	
	(1) Coef./s.e.	(2) Coef./s.e.	(1) Coef./s.e.	(2) Coef./s.e.	(1) Coef./s.e.	(2) Coef./s.e.
Female	-0.071***	-0.103***	-0.587***	-0.457***	-0.016	0.021
	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)
Household size	-0.039**	-0.037*	0.009	-0.038	-0.004	-0.011
	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)
Medium Income	0.092***	0.072***	-0.082**	-0.022	-0.019	-0.020
	(0.02)	(0.02)	(0.04)	(0.04)	(0.03)	(0.03)
High Income	0.163***	0.130***	-0.296***	-0.145***	0.022	-0.001
	(0.03)	(0.03)	(0.05)	(0.05)	(0.04)	(0.04)
Secondary Education	0.004	-0.016	-0.042	0.038	0.063**	0.057*
	(0.03)	(0.03)	(0.04)	(0.04)	(0.03)	(0.03)
Terciary Education	0.034	-0.007	-0.137***	-0.036	0.194***	0.152***
	(0.03)	(0.03)	(0.04)	(0.04)	(0.03)	(0.03)
Children at home	-0.044*	-0.046*	0.095**	0.019	-0.045	-0.034
	(0.03)	(0.03)	(0.04)	(0.04)	(0.03)	(0.03)
Out of labor force	-0.005	0.003	-0.030	-0.077	-0.069*	-0.065*
	(0.03)	(0.03)	(0.05)	(0.05)	(0.04)	(0.04)
Student	0.021	0.009	0.025	0.010	0.070	0.029
	(0.04)	(0.04)	(0.07)	(0.06)	(0.05)	(0.05)
Unemployed	-0.019	0.001	0.010	-0.039	0.025	0.026
	(0.04)	(0.04)	(0.06)	(0.05)	(0.04)	(0.04)
Self Employed	-0.017	-0.004	-0.213***	-0.199***	-0.048	-0.056
	(0.03)	(0.03)	(0.05)	(0.05)	(0.04)	(0.04)
Spain	-0.828***	-0.772***	1.156***	1.201***	0.254***	0.193***
	(0.03)	(0.03)	(0.05)	(0.05)	(0.04)	(0.04)
France	-0.132***	-0.133***	0.449***	0.297***	0.933***	0.928***
	(0.04)	(0.04)	(0.05)	(0.05)	(0.04)	(0.04)
Italy	-0.788***	-0.729***	0.919***	0.879***	-0.251***	-0.314***
	(0.03)	(0.03)	(0.05)	(0.05)	(0.03)	(0.03)
UK	-0.288***	-0.255***	0.137**	0.391***	0.022	0.082**
	(0.04)	(0.04)	(0.05)	(0.05)	(0.04)	(0.04)
Age: 16-25	0.024	-0.015	0.652***	0.633***	-0.155*	-0.159**
	(0.06)	(0.06)	(0.10)	(0.10)	(0.08)	(0.08)
Age: 26-30	0.111*	0.045	0.435***	0.542***	-0.279***	-0.236***
	(0.06)	(0.06)	(0.10)	(0.10)	(0.08)	(0.08)
Age: 31-40	0.102*	0.045	0.153*	0.427***	-0.300***	-0.210***
	(0.06)	(0.06)	(0.09)	(0.09)	(0.07)	(0.07)
Age: 41-50	0.088	0.042	-0.051	0.296***	-0.371***	-0.278***
	(0.06)	(0.06)	(0.09)	(0.09)	(0.07)	(0.07)
Age: 51-60	0.035	-0.002	-0.246***	0.219**	-0.352***	-0.243***
	(0.06)	(0.06)	(0.09)	(0.09)	(0.07)	(0.07)
Age: 61-75	0.006	-0.018	-0.678***	0.010	-0.417***	-0.292***
	(0.06)	(0.06)	(0.09)	(0.09)	(0.07)	(0.07)
Total online time	0.130***	0.000	0.238***	-0.279***	0.179***	0.020
	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)
Total (log of) clicks on:						
Other music websites	0.170***	0.145***	0.146***	0.079***	0.195***	0.158***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)

Continued on next page

TABLE 3.4: Determinants of music consumption, †

	Purchase		Downloading		Streaming	
	(1)	(2)	(1)	(2)	(1)	(2)
	Coef./s.e.	Coef./s.e.	Coef./s.e.	Coef./s.e.	Coef./s.e.	Coef./s.e.
Radio & music video websites	0.083***	0.067***	0.145***	0.027***	0.122***	0.083***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Ilegal Streaming websites	0.113***	0.098***	0.479***	0.357***	0.255***	0.232***
	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)
CD Purchase websites	0.232***	0.200***	0.319***	0.225***	0.286***	0.261***
	(0.03)	(0.03)	(0.05)	(0.04)	(0.03)	(0.03)
Online store websites		0.053***		0.066***		-0.017**
		(0.01)		(0.01)		(0.01)
Books websites		0.031***		0.006		0.044***
		(0.01)		(0.01)		(0.01)
Events websites		0.044***		-0.045***		0.038***
		(0.01)		(0.01)		(0.01)
Movies websites		0.026***		0.334***		0.088***
		(0.01)		(0.01)		(0.01)
Coupons websites		0.010		0.036***		-0.021***
		(0.01)		(0.01)		(0.01)
Instant messaging websites		0.017***		0.060***		0.043***
		(0.01)		(0.01)		(0.01)
Personal webpage websites		0.019**		0.267***		0.040***
		(0.01)		(0.01)		(0.01)
Global news websites		-0.005		-0.101***		0.034***
		(0.01)		(0.01)		(0.01)
Social Networks websites		0.006		0.006		0.025***
		(0.01)		(0.01)		(0.01)
Online gaming websites		-0.002		0.059***		-0.014***
		(0.00)		(0.01)		(0.00)
Constant	-0.694***	0.441***	-1.664***	2.076***	-1.703***	-0.469**
	(0.13)	(0.16)	(0.20)	(0.25)	(0.14)	(0.18)
Adjusted-R^2	0.185	0.198	0.237	0.334	0.225	0.243
No. of Obs.	16290	16290	16290	16290	16290	16290

† The dependent variable is the logarithm of the number of clicks on legal digital music purchase websites (first column), illegal digital music downloading websites (second column) and legal digital music streaming websites (third column). All regressors referring to clicks on a given type of website are in logarithm. Total time online is the logarithm of the total time spent online during the year, in seconds. Robust standard errors are in parenthesis. The reference country is Germany.

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

Table 3.5: Ordinary Least Squares (OLS) Results, Cross-Section data †

	(1) Coef./s.e.	(2) Coef./s.e.	(3) Coef./s.e.	(4) Coef./s.e.	(5) Coef./s.e.	(6) Coef./s.e.
Illegal download websites	0.064***	0.051***	0.044***	0.042***	0.039***	0.032***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Legal Streaming websites	0.140***	0.109***	0.097***	0.095***	0.090***	0.084***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Female	-0.044**	-0.053***	-0.056***	-0.056***	-0.047**	-0.087***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Household size	-0.030	-0.038**	-0.038**	-0.038**	-0.039**	-0.035*
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Medium Income	0.101***	0.093***	0.095***	0.096***	0.096***	0.074***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
High Income	0.175***	0.164***	0.172***	0.174***	0.173***	0.134***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Education	0.013	0.007	0.012	0.012	0.012	-0.008
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Children at home	-0.050*	-0.041	-0.046*	-0.047*	-0.044*	-0.044*
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Out of labor force	-0.012	0.003	0.010	0.010	0.003	0.011
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Student	0.030	0.016	0.009	0.009	0.013	0.006
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Unemployed	-0.036	-0.025	-0.022	-0.023	-0.021	-0.001
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Self Employed	-0.007	-0.006	-0.004	-0.004	-0.004	0.009
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Spain	-0.922***	-0.867***	-0.885***	-0.911***	-0.897***	-0.828***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
France	-0.283***	-0.218***	-0.246***	-0.244***	-0.234***	-0.221***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Italy	-0.823***	-0.774***	-0.809***	-0.811***	-0.805***	-0.734***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
UK	-0.267***	-0.269***	-0.289***	-0.287***	-0.297***	-0.273***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Age: 16-25	-0.002	0.004	0.015	0.013	0.009	-0.025
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
Age: 26-30	0.061	0.096	0.127**	0.124**	0.116*	0.045
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
Age: 31-40	0.038	0.089	0.137**	0.136**	0.119**	0.043
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
Age: 41-50	0.046	0.095*	0.142**	0.139**	0.120**	0.050
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
Age: 51-60	-0.033	0.037	0.096*	0.093	0.072	0.006
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
Age: 61-75	-0.038	0.024	0.090	0.087	0.066	-0.003
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
Total online time	0.193***	0.144***	0.107***	0.106***	0.104***	0.006
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Total (log of) clicks on:						
Other music websites		0.174***	0.156***	0.156***	0.147***	0.129***

Continued on next page

TABLE 3.5: Ordinary Least Squares (OLS) Results, Cross-Section data †

	(1) Coef./s.e.	(2) Coef./s.e.	(3) Coef./s.e.	(4) Coef./s.e.	(5) Coef./s.e.	(6) Coef./s.e.
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Radio & music video websites			0.069***	0.068***	0.067***	0.059***
			(0.01)	(0.01)	(0.01)	(0.01)
Ilegal Streaming websites				0.073***	0.071***	0.066***
				(0.02)	(0.02)	(0.02)
CD Purchase websites					0.194***	0.171***
					(0.03)	(0.03)
Online store websites						0.052***
						(0.01)
Books websites						0.026***
						(0.01)
Events websites						0.042***
						(0.01)
Movies websites						0.008
						(0.01)
Coupons websites						0.010*
						(0.01)
Instant messaging websites						0.012**
						(0.01)
Personal webpage websites						0.007
						(0.01)
Constant	-1.404***	-0.899***	-0.541***	-0.518***	-0.486***	0.446***
	(0.12)	(0.12)	(0.13)	(0.13)	(0.13)	(0.15)
Adjusted-R^2	0.162	0.186	0.192	0.193	0.197	0.207
No. of Obs.	16290	16290	16290	16290	16290	16290

† The dependent variable is the logarithm of the number of clicks on legal digital music purchase websites. All regressors referring to clicks on a given type of website are in logarithm. Total time online is the logarithm of the total time spent online during the year, in seconds. Robust standard errors are in parenthesis. The reference country is Germany.

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

Table 3.6: Tobit Results, Unconditional Marginal Effects , Cross-Section data †

	(1)	(2)	(3)	(4)	(5)	(6)
	Coef./s.e.	Coef./s.e.	Coef./s.e.	Coef./s.e.	Coef./s.e.	Coef./s.e.
Illegal download websites	0.047***	0.035***	0.028***	0.025***	0.022***	0.015***
	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
Legal Streaming websites	0.111***	0.081***	0.069***	0.067***	0.063***	0.056***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Female (d)	-0.050**	-0.059***	-0.061***	-0.061***	-0.054***	-0.098***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Household size	-0.034*	-0.042**	-0.042**	-0.041**	-0.042**	-0.037*
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Medium Income (d)	0.114***	0.105***	0.107***	0.108***	0.108***	0.084***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
High Income (d)	0.207***	0.192***	0.201***	0.203***	0.203***	0.154***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Education	0.026*	0.019	0.025*	0.025*	0.025*	0.004
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Children at home (d)	-0.057**	-0.049*	-0.054**	-0.056**	-0.053*	-0.056**
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Out of labor force (d)	-0.013	0.004	0.012	0.011	0.006	0.013
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Student (d)	0.039	0.022	0.015	0.013	0.017	0.015
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Unemployed (d)	-0.052	-0.040	-0.037	-0.039	-0.038	-0.014
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Self Employed (d)	-0.012	-0.011	-0.009	-0.010	-0.011	0.004
	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Spain (d)	-0.796***	-0.758***	-0.771***	-0.802***	-0.794***	-0.736***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
France (d)	-0.224***	-0.166***	-0.195***	-0.191***	-0.184***	-0.173***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Italy (d)	-0.708***	-0.675***	-0.701***	-0.703***	-0.700***	-0.641***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
UK (d)	-0.193***	-0.194***	-0.215***	-0.212***	-0.219***	-0.196***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Age: 16-25 (d)	-0.032	-0.021	-0.010	-0.012	-0.016	-0.061
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
Age: 26-30 (d)	0.013	0.051	0.086	0.082	0.075	-0.013
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
Age: 31-40 (d)	-0.018	0.037	0.089	0.086	0.073	-0.016
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
Age: 41-50 (d)	-0.009	0.043	0.093	0.090	0.074	-0.006
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
Age: 51-60 (d)	-0.103*	-0.034	0.028	0.022	0.006	-0.067
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
Age: 61-75 (d)	-0.083	-0.023	0.046	0.041	0.024	-0.049
	(0.06)	(0.06)	(0.07)	(0.07)	(0.06)	(0.06)
Total online time	0.237***	0.188***	0.147***	0.146***	0.145***	0.032**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Total (log of) clicks on:						

Continued on next page

Table 3.6: Tobit Results, Unconditional Marginal Effects , Cross-Section data †

	(1) Coef./s.e.	(2) Coef./s.e.	(3) Coef./s.e.	(4) Coef./s.e.	(5) Coef./s.e.	(6) Coef./s.e.
Other music websites		0.161***	0.142***	0.141***	0.135***	0.115***
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Radio & music video websites			0.072***	0.071***	0.070***	0.061***
			(0.01)	(0.01)	(0.01)	(0.01)
Ilegal Streaming websites				0.116***	0.114***	0.110***
				(0.02)	(0.02)	(0.02)
CD Purchase websites					0.140***	0.116***
					(0.02)	(0.02)
Online store websites						0.069***
						(0.01)
Books websites						0.021***
						(0.01)
Events websites						0.045***
						(0.01)
Movies websites						0.015*
						(0.01)
Coupons websites						0.006
						(0.01)
Instant messaging websites						0.014***
						(0.01)
Personal webpage websites						0.006
						(0.01)
No. of Obs.	16290	16290	16290	16290	16290	16290

† The table presents unconditional marginal effects. The dependent variable is the logarithm of the number of clicks on legal digital music purchase websites. All regressors referring to clicks on a given type of website are in logarithm. Total time online is the logarithm of the total time spent online during the year, in seconds. (d)= dummy variable. Robust standard errors are in parenthesis. The reference country is Germany.

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

Table 3.7: Ordinary Least Squares (OLS) Results, monthly data †

	(1) Coef./s.e.	(2) Coef./s.e.	(3) Coef./s.e.	(4) Coef./s.e.	(5) Coef./s.e.	(6) Coef./s.e.
Illegal download websites	0.045***	0.033***	0.031***	0.031***	0.030***	0.026***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Legal Streaming websites	0.105***	0.078***	0.075***	0.075***	0.074***	0.071***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Female	-0.016***	-0.019***	-0.019***	-0.019***	-0.018***	-0.024***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Household size	-0.003	-0.004	-0.004	-0.004	-0.004	-0.003
	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Medium Income	0.008	0.007	0.007	0.007	0.007	0.007
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
High Income	0.020**	0.019**	0.020**	0.020**	0.020**	0.020**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Education	-0.000	-0.002	-0.001	-0.001	-0.001	-0.003
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Children at home	-0.014**	-0.010	-0.011	-0.011	-0.010	-0.009
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Out of labor force	0.005	0.008	0.008	0.008	0.006	0.004
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Student	0.007	-0.004	-0.006	-0.006	-0.005	-0.008
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Unemployed	-0.001	-0.001	-0.002	-0.002	-0.002	-0.003
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Self Employed	0.003	0.002	0.000	0.000	-0.000	0.001
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
January	0.040***	0.039***	0.041***	0.041***	0.042***	0.035***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
February	0.014***	0.013**	0.014***	0.015***	0.015***	0.012**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
March	0.020***	0.020***	0.021***	0.021***	0.022***	0.018***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
April	0.004	0.006	0.007	0.007	0.007	0.006
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
May	-0.003	-0.003	-0.002	-0.002	-0.002	-0.003
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
June	0.007	0.007	0.007	0.007	0.007	0.006
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
August	-0.011**	-0.008*	-0.008*	-0.008*	-0.008*	-0.009*
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
September	-0.017***	-0.014***	-0.013***	-0.013***	-0.013***	-0.015***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
October	0.005	0.007	0.008	0.008	0.008	0.003
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
November	-0.007	-0.008	-0.008	-0.008	-0.008	-0.014***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
December	0.019***	0.019***	0.019***	0.019***	0.019***	0.012**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Age: 16-25	0.015	0.012	0.011	0.011	0.010	0.004
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Age: 26-30	0.037**	0.048***	0.051***	0.051***	0.049***	0.036**

Continued on next page

TABLE 3.7: Ordinary Least Squares (OLS) Results, monthly data †

	(1) Coef./s.e.	(2) Coef./s.e.	(3) Coef./s.e.	(4) Coef./s.e.	(5) Coef./s.e.	(6) Coef./s.e.
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Age: 31-40	0.020	0.038***	0.043***	0.043***	0.041***	0.026**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Age: 41-50	0.020	0.038***	0.042***	0.042***	0.039***	0.026**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Age: 51-60	-0.001	0.023*	0.030**	0.030**	0.027**	0.015
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Age: 61-75	-0.002	0.025*	0.034**	0.034**	0.030**	0.020
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Total online time	0.022***	0.013***	0.010***	0.010***	0.010***	0.001
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Total (log of) clicks on:						
Other music websites		0.082***	0.075***	0.075***	0.073***	0.065***
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Radio & music video websites			0.026***	0.026***	0.026***	0.021***
			(0.00)	(0.00)	(0.00)	(0.00)
Ilegal Streaming websites				0.007	0.006	0.004
				(0.01)	(0.01)	(0.01)
CD Purchase websites					0.174***	0.166***
					(0.04)	(0.04)
Online store websites						0.012***
						(0.00)
Books websites						0.014***
						(0.00)
Events websites						0.007***
						(0.00)
Movies websites						0.005***
						(0.00)
Coupons websites						0.004***
						(0.00)
Instant messaging websites						0.006***
						(0.00)
Personal webpage websites						0.004*
						(0.00)
Constant	-0.198***	-0.152***	-0.145***	-0.145***	-0.142***	-0.073***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Adjusted-R^2	0.059	0.081	0.083	0.083	0.085	0.089
No. of Obs.	195478	195478	195478	195478	195478	195478

† The dependent variable is the logarithm of the number of clicks on legal digital music purchase websites. All regressors referring to clicks on a given type of website are in logarithm. Total time online is the logarithm of the total time spent online during the month, in seconds. Standard errors are in parenthesis and clustered at the individual level. All specifications include regional dummies. The reference month is July.

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

Table 3.8: Tobit Results, Unconditional Marginal Effects monthly data †

	(1)	(2)	(3)	(4)	(5)	(6)
	Coef./s.e.	Coef./s.e.	Coef./s.e.	Coef./s.e.	Coef./s.e.	Coef./s.e.
Illegal download websites	0.024***	0.018***	0.017***	0.017***	0.017***	0.014***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Legal Streaming websites	0.040***	0.027***	0.025***	0.025***	0.025***	0.024***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Female (d)	-0.014***	-0.016***	-0.016***	-0.016***	-0.015***	-0.019***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Household size	-0.001	-0.002	-0.002	-0.002	-0.002	-0.002
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Medium Income (d)	0.012***	0.010**	0.010**	0.010**	0.010**	0.009**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
High Income (d)	0.030***	0.025***	0.026***	0.026***	0.026***	0.024***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Education	0.005**	0.003	0.004*	0.004*	0.004*	0.003
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Children at home (d)	-0.010**	-0.008*	-0.009*	-0.009*	-0.009*	-0.010**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Out of labor force (d)	-0.003	0.001	0.001	0.001	0.001	-0.000
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Student (d)	0.007	-0.001	-0.002	-0.002	-0.002	-0.003
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Unemployed (d)	-0.007	-0.006	-0.006	-0.006	-0.006	-0.006
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Self Employed (d)	-0.005	-0.005	-0.006	-0.006	-0.006	-0.005
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
January (d)	0.030***	0.031***	0.032***	0.033***	0.033***	0.030***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
February (d)	0.010**	0.009**	0.011***	0.011***	0.011***	0.010**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
March (d)	0.013***	0.013***	0.015***	0.015***	0.015***	0.014***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
April (d)	0.003	0.004	0.005	0.005	0.005	0.005
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
May (d)	-0.004	-0.003	-0.002	-0.002	-0.002	-0.003
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
June (d)	0.004	0.004	0.004	0.004	0.004	0.004
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
August (d)	-0.012***	-0.010***	-0.010***	-0.009***	-0.010***	-0.010***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
September (d)	-0.016***	-0.014***	-0.014***	-0.014***	-0.014***	-0.015***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
October (d)	0.005	0.007*	0.008*	0.008**	0.008**	0.005
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
November (d)	-0.007*	-0.007*	-0.006	-0.006	-0.006	-0.010***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
December (d)	0.020***	0.021***	0.022***	0.022***	0.022***	0.016***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Age: 16-25 (d)	-0.000	0.003	0.004	0.004	0.003	-0.000
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)

Continued on next page

Table 3.8: Tobit Results, Unconditional Marginal Effects monthly data †

	(1)	(2)	(3)	(4)	(5)	(6)
	Coef./s.e.	Coef./s.e.	Coef./s.e.	Coef./s.e.	Coef./s.e.	Coef./s.e.
Age: 26-30 (d)	0.007	0.020*	0.024*	0.024*	0.023*	0.016
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Age: 31-40 (d)	-0.007	0.010	0.015	0.015	0.014	0.008
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Age: 41-50 (d)	-0.007	0.009	0.014	0.014	0.012	0.008
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Age: 51-60 (d)	-0.018*	0.001	0.007	0.006	0.005	0.002
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Age: 61-75 (d)	-0.014	0.007	0.013	0.013	0.011	0.010
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Total online time	0.045***	0.033***	0.030***	0.030***	0.030***	0.019***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Total (log of) clicks on:						
Other music websites		0.041***	0.038***	0.038***	0.037***	0.034***
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Radio & music video websites			0.013***	0.013***	0.013***	0.010***
			(0.00)	(0.00)	(0.00)	(0.00)
Ilegal Streaming websites				0.013**	0.013**	0.013**
				(0.01)	(0.01)	(0.01)
CD Purchase websites					0.036***	0.033***
					(0.01)	(0.01)
Online store websites						0.009***
						(0.00)
Books websites						0.005***
						(0.00)
Events websites						0.001
						(0.00)
Movies websites						0.007***
						(0.00)
Coupons websites						0.001
						(0.00)
Instant messaging websites						0.002
						(0.00)
Personal webpage websites						0.000
						(0.00)
Constant						
sigma						
Constant						
No. of Obs.	195478	195478	195478	195478	195478	195478

† The table presents unconditional marginal effects. The dependent variable is the logarithm of the number of clicks on legal digital music purchase websites. All regressors referring to clicks on a given type of website are in logarithm. Total time online is the logarithm of the total time spent online during the month, in seconds. (d)= dummy variable. Standard errors are in parenthesis and clustered at the individual level. All specifications include regional dummies. The reference month is July.

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

TABLE 3.9: Fixed Effects Estimation †

	(1) Coef./s.e.	(2) Coef./s.e.	(3) Coef./s.e.	(4) Coef./s.e.	(5) Coef./s.e.	(6) Coef./s.e.
Illegal download websites	0.036***	0.029***	0.027***	0.027***	0.027***	0.022***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Legal Streaming websites	0.067***	0.055***	0.052***	0.052***	0.051***	0.049***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Total online time	0.017***	0.013***	0.011***	0.011***	0.011***	0.003***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Total (log of) clicks on:						
Other music websites		0.061***	0.057***	0.057***	0.056***	0.051***
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Radio & music video websites			0.030***	0.030***	0.029***	0.024***
			(0.00)	(0.00)	(0.00)	(0.00)
Ilegal Streaming websites				0.018**	0.017*	0.016*
				(0.01)	(0.01)	(0.01)
CD Purchase websites					0.103***	0.100***
					(0.02)	(0.02)
Online store websites						0.008***
						(0.00)
Books websites						0.008***
						(0.00)
Events websites						0.003*
						(0.00)
Movies websites						0.010***
						(0.00)
Coupons websites						0.007***
						(0.00)
Instant messaging websites						0.011***
						(0.00)
Personal webpage websites						0.003**
						(0.00)
Constant	-0.052***	-0.038***	-0.033***	-0.033***	-0.033***	-0.015**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Adjusted-R^2	0.018	0.029	0.031	0.031	0.032	0.034
No. of Obs.	195478	195478	195478	195478	195478	195478

† The dependent variable is the logarithm of the number of clicks on legal digital music purchase websites. All regressors referring to clicks on a given type of website are in logarithm. Total time online is the logarithm of the total time spent online during the month, in seconds. All specifications include monthly fixed effects. Standard errors are in parenthesis and clustered at the individual level.

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

Table 3.10: Fixed Effects Results, country interactions †

	(1) Coef./s.e.	(2) Coef./s.e.	(3) Coef./s.e.	(4) Coef./s.e.	(5) Coef./s.e.	(6) Coef./s.e.
Illegal download websites	0.053***	0.045***	0.043***	0.043***	0.042***	0.038***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Illegal download×Spain	-0.030***	-0.029***	-0.028***	-0.029***	-0.029***	-0.030***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Illegal download×France	0.004	0.007	0.007	0.007	0.008	0.007
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Illegal download×Italy	-0.031***	-0.030***	-0.029***	-0.029***	-0.029***	-0.029***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Illegal download×UK	-0.012	-0.011	-0.011	-0.011	-0.012	-0.011
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Legal Streaming websites	0.081***	0.068***	0.066***	0.066***	0.066***	0.062***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Legal streaming×Spain	-0.038***	-0.039***	-0.040***	-0.041***	-0.040***	-0.039***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Legal streaming×France	0.004	0.006	0.005	0.005	0.005	0.006
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Legal streaming×Italy	-0.027*	-0.027*	-0.028**	-0.029**	-0.029**	-0.029**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Legal streaming×UK	-0.018	-0.017	-0.018	-0.018	-0.019	-0.018
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Total online time	0.017***	0.013***	0.011***	0.011***	0.011***	0.003***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Total (log of) clicks on:						
Other music websites		0.061***	0.057***	0.057***	0.056***	0.051***
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Radio & music video websites			0.030***	0.029***	0.029***	0.024***
			(0.00)	(0.00)	(0.00)	(0.00)
Ilegal Streaming websites				0.026***	0.026***	0.024***
				(0.01)	(0.01)	(0.01)
CD Purchase websites					0.102***	0.099***
					(0.02)	(0.02)
Online store websites						0.008***
						(0.00)
Books websites						0.009***
						(0.00)
Events websites						0.003*
						(0.00)
Movies websites						0.010***
						(0.00)
Coupons websites						0.007***
						(0.00)
Instant messaging websites						0.011***
						(0.00)
Personal webpage websites						0.003**
						(0.00)
Constant	-0.052***	-0.037***	-0.033***	-0.033***	-0.033***	-0.014**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Adjusted-R^2	0.019	0.030	0.032	0.032	0.033	0.035

Continued on next page

Table 3.10: Fixed Effects Results, country interactions †

	(1) Coef./s.e.	(2) Coef./s.e.	(3) Coef./s.e.	(4) Coef./s.e.	(5) Coef./s.e.	(6) Coef./s.e.
No. of Obs.	195478	195478	195478	195478	195478	195478

† The dependent variable is the logarithm of the number of clicks on legal digital music purchase websites. All regressors referring to clicks on a given type of website are in logarithm. Total time online is the logarithm of the total time spent online during the month, in seconds. All specifications include monthly fixed effects. Standard errors are in parenthesis and clustered at the individual level.

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

Table 3.11: Fixed Effects Results, interaction with music interest †

	(1) Coef./s.e.	(2) Coef./s.e.	(3) Coef./s.e.	(4) Coef./s.e.
Illegal download websites	0.022***	0.018***	0.013***	0.012***
	(0.00)	(0.00)	(0.00)	(0.00)
Illegal download x Other music websites			0.009***	0.008***
			(0.00)	(0.00)
Illegal download x Radio & music videos		0.003***		0.001
		(0.00)		(0.00)
Legal Streaming websites	0.049***	0.033***	0.027***	0.020***
	(0.00)	(0.00)	(0.00)	(0.00)
Legal streaming x Other music websites			0.014***	0.012***
			(0.00)	(0.00)
Legal streaming x Radio & music videos		0.012***		0.008***
		(0.00)		(0.00)
Total online time	0.003***	0.003***	0.004***	0.004***
	(0.00)	(0.00)	(0.00)	(0.00)
Total (log of) clicks on:				
Other music websites	0.051***	0.051***	0.037***	0.038***
	(0.00)	(0.00)	(0.00)	(0.00)
Radio & music video websites	0.024***	0.016***	0.023***	0.020***
	(0.00)	(0.00)	(0.00)	(0.00)
Ilegal Streaming websites	0.016*	0.014	0.011	0.011
	(0.01)	(0.01)	(0.01)	(0.01)
CD Purchase websites	0.100***	0.098***	0.092***	0.092***
	(0.02)	(0.02)	(0.02)	(0.02)
Online store websites	0.008***	0.008***	0.008***	0.008***
	(0.00)	(0.00)	(0.00)	(0.00)
Books websites	0.008***	0.008***	0.008***	0.008***
	(0.00)	(0.00)	(0.00)	(0.00)
Events websites	0.003*	0.003	0.003	0.003
	(0.00)	(0.00)	(0.00)	(0.00)
Movies websites	0.010***	0.011***	0.012***	0.012***
	(0.00)	(0.00)	(0.00)	(0.00)
Coupons websites	0.007***	0.007***	0.007***	0.007***
	(0.00)	(0.00)	(0.00)	(0.00)
Instant messaging websites	0.011***	0.011***	0.011***	0.011***
	(0.00)	(0.00)	(0.00)	(0.00)
Personal webpage websites	0.003**	0.003**	0.003*	0.003*
	(0.00)	(0.00)	(0.00)	(0.00)
Constant	-0.015**	-0.016**	-0.017***	-0.017***
	(0.01)	(0.01)	(0.01)	(0.01)
Adjusted-R^2	0.034	0.035	0.036	0.036
No. of Obs.	195478	195478	195478	195478

† The dependent variable is the logarithm of the number of clicks on legal digital music purchase websites. All regressors referring to clicks on a given type of website are in logarithm. Total time online is the logarithm of the total time spent online during the month, in seconds. All specifications include monthly fixed effects. Standard errors are in parenthesis and clustered at the individual level.

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

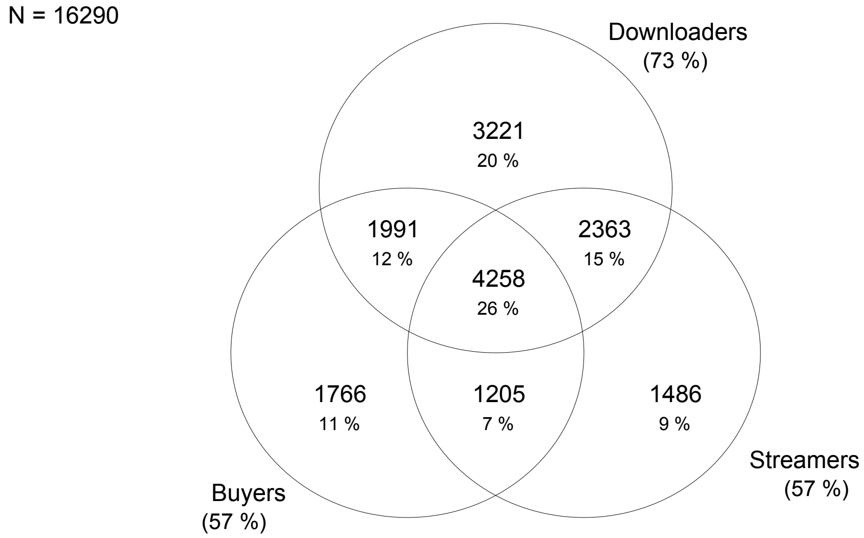

FIGURE 3.1: Composition of the sample by types of music consumer.

Bibliography

Artis, M. and F. Nixson (2001): *The Economics of the European Union: Policy and Analysis*, New York: Oxford University Press, third ed.

Bai, J. and J. Waldfogel (2012): "Movie piracy and sales displacement in two samples of Chinese consumers," *Information Economics and Policy*, 24, 187–196.

Barajas, A., E. Huergo, and L. Moreno (2011): "Measuring the economic impact of research joint ventures supported by the EU Framework Programme," *The Journal of Technology Transfer*, 1–26.

Barbosa, K. (2009): "Free-Riding in Procurement Design," LERNA Working Papers 09.29.305, LERNA, University of Toulouse.

Bastard, I., M. Bourreau, and F. Moreau (2012): "L'impact du Piratage sur l'Achat et le Téléchargement Légal: une Comparaison de Quatre Filières Culturelles," Working paper.

Belderbos, R., M. Carree, and B. Lokshin (2004): "Cooperative R&D and firm performance," *Research Policy*, 33, 1477–1492.

Belleflamme, P. and M. Peitz (2010): "Digital Piracy: Theory," CESifo Working Paper Series 3222, CESifo Group Munich.

Benfratello, L. and A. Sembenelli (2002): "Research joint ventures and firm level performance," *Research Policy*, 31, 493–507.

Bernstein, J. I. and M. I. Nadiri (1989): "Research and Development and Intra-industry Spillovers: An Empirical Application of Dynamic Duality," *Review of Economic Studies*, 56, 249–67.

Bloom, N., M. Schankerman, and J. Van Reenen (2013): "Identifying Technology Spillovers and Product Market Rivalry," *Econometrica*, 81, 1347–1393.

Branstetter, L. and M. Sakakibara (1998): "Japanese Research Consortia: A Microeconometric Analysis of Industrial Policy," *Journal of Industrial Economics*, 46, 207–33.

——— (2002): "When Do Research Consortia Work Well and Why? Evidence from Japanese Panel Data," *American Economic Review*, 92, 143–159.

Brickley, J. A. (1999): "Incentive Conflicts and Contractual Restraints: Evidence from Franchising," *Journal of Law and Economics*, 42, 745–74.

Cassiman, B. and R. Veugelers (2002): "R&D Cooperation and Spillovers: Some Empirical Evidence from Belgium," *American Economic Review*, 92, 1169–1184.

Cecere, G., N. Corrocher, and F. Scarica (2012): "Why do pirates buy music online? An empirical analysis on a sample of college students," *Economics Bulletin*, 32, 2955–2968.

Cincera, M., L. Kempen, B. van Pottelsberghe de la Potterie, R. Veugelers, and C. Villegas Sanchez (2003): "Productivity growth, R&D and the role of international collaborative agreements: Some evidence for Belgian manufacturing companies," *Brussels Economic Review/Cahiers Economiques de Bruxelles*, 46, 107–140.

Cohen, W. M. and D. A. Levinthal (1989): "Innovation and Learning: The Two Faces of R&D," *Economic Journal*, 99, 569–96.

Czarnitzki, D., B. Ebersberger, and A. H. Fier (2007): "The Relationship between R&D Collaboration, Subsidies and R&D Performance: Empirical Evidence from Finland and Germany," *Journal of Applied Econometrics*, 22, 1347–1366.

Danaher, B., S. Dhanasobhon, M. D. Smith, and R. Telang (2010): "Converting Pirates Without Cannibalizing Purchasers: The Impact of Digital Distribution on Physical Sales and Internet Piracy," *Marketing Science*, 29, 1138–1151.

Danaher, B. and M. D. Smith (2013): "Gone in 60 Seconds: The Impact of the Megaupload Shutdown on Movie Sales," Working Paper 16882.

Danaher, B., M. D. Smith, R. Telang, and S. Chen (2012): "The Effect of Graduated Response Anti-Piracy Laws on Music Sales: Evidence from an Event Study in France," Working paper.

DangNguyen, G., S. Dejean, and F. Moreau (2012): "Are Streaming and Other Music Consumption Modes Substitutes or Complements?" Working paper.

Darr, E. D., L. Argote, and D. Epple (1995): "The Acquisition, Transfer, and Depreciation of Knowledge in Service Organizations: Productivity in Franchises," *Management Science*, 41, 1750–1762.

De Bondt, R. (1997): "Spillovers and innovative activities," *International Journal of Industrial Organization*, 15, 1–28.

de Comptes, C. (2005): "Les transports publics urbains: rapport au président de la république suivi des réponses des administrations et des organismes intéressés." Tech. rep.

Dekker, R. and A. Kleinknecht (2008): "The EU Framework Programs: Are they worth doing?" MPRA Paper 8503, University Library of Munich, Germany.

Domenach, O. (1987): "Métamorphose des Rapports Contractuels dans les Transports Publics Urbains," *Transports*, 326.

Einiö, E. (2012): "The Effects of Government R&D Subsidies on Company Performance: Evidence from the ERDF Population-Density Rule," forthcoming, *The Review of Economics and Statistics.*

European Commission (2000): *Participating in the European Research Programmes: Guide for Applicants under the Fifth Framework Programme for European Research and Technological Development (1999-2002).*, Silicon Valley, silver ed.

——— (2001): *Guide for Proposers: Information Society Technologies (IST) Programme, Part 1*, third edition ed.

Fisher, R., W. Polt, and N. Vonortas (2009): "The impact of publicly funded research on innovation: An analysis of European Framework Programmes for Research and Development," *Report for the European Commission (DG Enterprise and Industry)*, Brussels.

Gagnepain, P. and M. Ivaldi (2002): "Incentive Regulatory Policies: The Case of Public Transit Systems in France," *RAND Journal of Economics*, 33, 605–629.

——— (2010): "Contract Choice, Incentives, and Political Capture in the Public Sector," CEPR Discussion Papers 8053, C.E.P.R. Discussion Papers.

Gagnepain, P., M. Ivaldi, and D. Martimort (2013): "The Cost of Contract Renegotiation: Evidence from the Local Public Sector," *American Economic Review*, 103, 2352–83.

Gelabert, L., A. Fosfuri, and J. A. Tribó (2009): "Does The Effect Of Public Support For R&D Depend On The Degree Of Appropriability?" *The Journal of Industrial Economics*, 57, 736–767.

Georghiou, L. (2001): "Evolving frameworks for European collaboration in research and technology," *Research Policy*, 30, 891–903.

Griliches, Z. (1992): "The Search for R&D Spillovers," *Scandinavian Journal of Economics*, 94, S29–47.

Hagedoorn, J., A. N. Link, and N. S. Vonortas (2000): "Research partnerships," *Research Policy*, 29, 567–586.

Hammond, R. G. (2012): "Profit Leak? Pre-Release File Sharing and the Music Industry," Working paper, North Carolina State University, Raleigh, North Carolina.

Handke, C. (2012): "Digital copying and the supply of sound recordings," *Information Economics and Policy*, 24, 15 – 29, ¡ce:title¿The Economics of Digital Media Markets¡/ce:title¿.

Hernán, R., P. L. Marín, and G. Siotis (2003): "An empirical evaluation of the determinants of Research Joint Venture Formation," *Journal of Industrial Economics*, 51, 75–89.

Hui, K.-L. and I. Png (2003): "Piracy and the Legitimate Demand for Recorded Music," *The B.E. Journal of Economic Analysis & Policy*, 11.

IDEI (1999): "Network Industries and Public Service." Tech. rep.

IFPI (2011): "International Federation of the Phonographic Industry Digital Music Report 2011," http://www.ifpi.org/content/library/dmr2011.pdf.

——— (2012): "International Federation of the Phonographic Industry Digital Music Report 2012," http://www.ifpi.org/content/library/dmr2012.pdf.

Imbens, G. W. and J. D. Angrist (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467–75.

Klette, T. J. (1996): "R&D, Scope Economies, and Plant Performance," *RAND Journal of Economics*, 27, 502–522.

Klette, T. J. and J. Moen (1999): "From Growth Theory to Technology Policy - Coordination Problems in Theory and Practice," *Nordic Journal of Political Economy*, 25, 53–74.

Klette, T. J., J. Moen, and Z. Griliches (2000): "Do subsidies to commercial R&D reduce market failures? Microeconometric evaluation studies," *Research Policy*, 29, 471–495.

Laffont, J. and J. Tirole (1993): *A Theory of Incentives in Procurement and Regulation*, Cambridge: MIT Press.

Lafontaine, F. and M. Slade (2007): "Vertical Integration and Firm Boundaries: The Evidence," *Journal of Economic Literature*, 45, 629–685.

Lafontaine, F. and M. E. Slade (1997): "Retail Contracting: Theory and Practice," *Journal of Industrial Economics*, 45, 1–25.

Landes, W. M. and R. A. Posner (1989): "An Economic Analysis of Copyright Law," *The Journal of Legal Studies*, 18, 325–63.

LAREDO, P. (1998): "The networks promoted by the framework programme and the questions they raise about its formulation and implementation," *Research Policy*, 27, 589–598.

LEVINTHAL, D. A. AND J. G. MARCH (1993): "The myopia of learning," *Strategic Management Journal*, 14, 95–112.

LICHTENBERG, F. R. (1988): "The Private R&D Investment Response to Federal Design and Technical Competitions," *American Economic Review*, 78, 550–59.

LIEBOWITZ, S. J. (2008): "Research Note–Testing File Sharing's Impact on Music Album Sales in Cities," *Management Science*, 54, 852–859.

LIMAYEM, M., M. KHALIFA, AND W. W. CHIN (2004): "Factors motivating software piracy: a longitudinal study," *Engineering Management, IEEE Transactions on*, 51, 414–425.

LUUKKONEN, T. (1998): "The difficulties in assessing the impact of EU framework programmes," *Research Policy*, 27, 599–610.

MANDEL, P. AND B. SÜSSMUTH (2012): "Determinants of Digital Piracy: A Re-examination of Results," *Journal of Economics and Statistics (Jahrbuecher fuer Nationaloekonomie und Statistik)*, 232, 394–413.

MARÍN, P. L. AND G. SIOTIS (2008): "Public policies towards Research Joint Venture: Institutional design and participants' characteristics," *Research Policy*, 37, 1057–1065.

MARRON, D. AND D. STEEL (2000): "Which countries protect intellectual property? The case of software piracy." *Economic Inquiry*, 38, 159174.

OBERHOLZER-GEE, F. AND K. STRUMPF (2007): "The Effect of File Sharing on Record Sales: An Empirical Analysis," *Journal of Political Economy*, 115, 1 – 42.

——— (2010): "File Sharing and Copyright," in *Innovation Policy and the Economy, Volume 10*, National Bureau of Economic Research, Inc, NBER Chapters, 19–55.

PEITZ, M. AND P. WAELBROECK (2004): "The Effect of Internet Piracy on Music Sales: Cross-Section Evidence," *Review of Economic Research on Copyright Issues*, 1, 71–79.

——— (2006): "Why the music industry may gain from free downloading – The role of sampling," *International Journal of Industrial Organization*, 24, 907–913.

RAMAYAH, T., N. H. AHMAD, L. G. CHIN, AND M. LO (2009): "Testing a Causal Model of Internet Piracy Behavior Among University Students," *European Journal of Scientific Research*, 29, 206–214.

RAMAYAH, T., H. RINNE, AND A. KAU (1990): "The morality of software piracy: A cross-cultural analysis," *Journal of Business Ethics*, 9, 655–664.

Rob, R. and J. Waldfogel (2006): "Piracy on the High C's: Music Downloading, Sales Displacement, and Social Welfare in a Sample of College Students," *Journal of Law and Economics*, 49, 29–62.

——— (2007): "Piracy on the Silver Screen," *Journal of Industrial Economics*, 55, 379–395.

Röller, L.-H., R. Siebert, and M. M. Tombak (2007): "Why Firms Form (or do not Form) RJVS," *Economic Journal*, 117, 1122–1144.

Sakakibara, M. (2001): "The Diversity of R&D Consortia and Firm Behavior: Evidence from Japanese Data," *Journal of Industrial Economics*, 49, 181–96.

Shapiro, C. and H. R. Varian (1999): *Information rules: a strategic guide to the network economy*, Harvard Business School Press.

Siebert, R. (1996): "The Impact of Research Joint Ventures on Firm Performance: An Empirical Assessment," CIG Working Papers FS IV 96-03, Wissenschaftszentrum Berlin (WZB), Research Unit: Competition and Innovation (CIG).

Sims, R., H. Cheng, and H. Teegen (1996): "Toward a profile of student software piraters," *Journal of Business Ethics*, 15, 839–849.

Smith, M. D. and R. Telang (2012): "Assessing the Academic Literature Regarding the Impact of Media Piracy on Sales," Working paper.

Szulanski, G. (1996): "Exploring Internal Stickiness: Impediments to the Transfer of Best Practice within the Firm." *Strategic Management Journal*, 17, 27–43.

Waldfogel, J. (2010): "Music file sharing and sales displacement in the iTunes era," *Information Economics and Policy*, 22, 306–314.

——— (2011): "Bye, Bye, Miss American Pie? The Supply of New Recorded Music Since Napster," NBER Working Papers 16882, National Bureau of Economic Research, Inc.

——— (2012a): "And the Bands Played on: Digital Disintermediation and the Quality of New Recorded Music," Working paper.

——— (2012b): "Copyright Protection, Technological Change, and the Quality of New Products: Evidence from Recorded Music since Napster," *Journal of Law and Economics*, 55, 715 – 740.

——— (2012c): "Copyright Research in the Digital Age: Moving from Piracy to the Supply of New Products," *American Economic Review*, 102, 337–42.

Waldman, M. (2013): "Piracy of Intellectual property: Past, Present and Future," Paper presented at the 2013 annual congress of the Society for Economic Research on Copyright Issues. Available at http://www.serci.org/2013/waldman.pdf.

Wallsten, S. J. (2000): "The Effects of Government-Industry R&D Programs on Private R&D: The Case of the Small Business Innovation Research Program," *RAND Journal of Economics*, 31, 82–100.

Zentner, A. (2006): "Measuring the Effect of File Sharing on Music Purchases," *Journal of Law and Economics*, 49, 63–90.

——— (2009): "Ten Years of File Sharing and Its Effect on International Physical and Digital Music Sales," Working paper, University of Dallas, Dallas, Texas.

Zobel, R. A. (1999): "The European Union Fifth Framework Programme," *Portland International Conference on Management of Engineering and Technology*, 2, 571–576.